GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 02:58:27 ; Search time 256 Seconds
(without alignments)
9299.265 Million cell updates/sec

Title: US-09-998-284-1
Perfect score: 1644
Sequence: 1 atggctactttgccacaaaa.....ctaagcagactaaatagtag 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
	1	1644	100.0	1644	10	US-09-998-284-1	Sequence 1, Appli
	2	1279.2	77.8	1801	10	US-09-824-053-30	Sequence 30, Appl
c	3	42.6	2.6	647	10	US-09-770-149-554	Sequence 554, App
	4	41.8	2.5	1584	9	US-09-938-842A-1942	Sequence 1942, Ap
	5	41.8	2.5	2397	9	US-10-170-656-3	Sequence 3, Appli
	6	41.4	2.5	1620	9	US-09-938-842A-2266	Sequence 2266, Ap
	7	40.2	2.4	1599	9	US-09-938-842A-281	Sequence 281, App
	8	39.2	2.4	1757	9	US-09-258-031B-70	Sequence 70, Appl
	9	38.6	2.3	1527	9	US-09-258-031B-72	Sequence 72, Appl
	10	38.6	2.3	1584	9	US-09-938-842A-1765	Sequence 1765, Ap
	11	37.2	2.3	1599	9	US-09-938-842A-1762	Sequence 1762, Ap
c	12	36.4	2.2	861	10	US-09-801-368-187	Sequence 187, App
	13	36.4	2.2	1589	9	US-09-258-031B-19	Sequence 19, Appl
	14	36.4	2.2	1784	9	US-09-258-031B-15	Sequence 15, Appl
	15	35.8	2.2	941	9	US-10-123-155-464	Sequence 464, App
	16	35.8	2.2	7766	9	US-10-222-162-3	Sequence 3, Appli
	17	35.8	2.2	7766	9	US-10-143-024-3	Sequence 3, Appli
c	18	35.4	2.2	513509	9	US-09-754-853A-4	Sequence 4, Appli
c	19	35	2.1	1756	9	US-10-114-043-7	Sequence 7, Appli
c	20	35	2.1	2306	9	US-10-114-043-1	Sequence 1, Appli
	21	34.8	2.1	292	10	US-09-924-035A-19	Sequence 19, Appl
c	22	34.4	2.1	604	10	US-09-770-149-927	Sequence 927, App
	23	34.2	2.1	261	9	US-09-258-031B-45	Sequence 45, Appl
	24	34.2	2.1	1830121	9	US-10-329-960-1	Sequence 1, Appli
	25	34	2.1	377	9	US-09-258-031B-42	Sequence 42, Appl
	26	34	2.1	377	9	US-09-258-031B-43	Sequence 43, Appl
	27	34	2.1	408	10	US-09-878-574-18	Sequence 18, Appl
	28	34	2.1	695	9	US-09-258-031B-24	Sequence 24, Appl
	29	34	2.1	7038	9	US-10-239-676-203	Sequence 203, App
	30	33.8	2.1	1981	9	US-09-258-031B-57	Sequence 57, Appl
	31	33.6	2.0	1593	9	US-09-938-842A-319	Sequence 319, App
c	32	33.4	2.0	596	10	US-09-864-761-8758	Sequence 8758, Ap
	33	33.4	2.0	1887	9	US-09-738-626-2031	Sequence 2031, Ap
c	34	33.4	2.0	3309400	9	US-09-738-626-1	Sequence 1, Appli
	35	33.2	2.0	1458	9	US-10-047-412A-3	Sequence 3, Appli
	36	33.2	2.0	1536	9	US-09-938-842A-222	Sequence 222, App
c	37	33	2.0	490	12	US-10-071-751-41	Sequence 41, Appl
	38	33	2.0	3172	9	US-09-954-531-158	Sequence 158, App
	39	33	2.0	3172	9	US-09-954-531-580	Sequence 580, App
	40	32.8	2.0	416	9	US-09-764-868-173	Sequence 173, App
	41	32.8	2.0	716	9	US-10-123-155-512	Sequence 512, App
	42	32.8	2.0	993	9	US-09-738-626-1095	Sequence 1095, Ap
	43	32.8	2.0	3309400	9	US-09-738-626-1	Sequence 1, Appli
	44	32.6	2.0	1530	9	US-09-258-031B-74	Sequence 74, Appl
	45	32.6	2.0	1626	9	US-09-938-842A-2575	Sequence 2575, Ap

ALIGNMENTS

RESULT 1
US-09-998-284-1
; Sequence 1, Application US/09998284
; Patent No. US20020106361A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, et al.
; TITLE OF INVENTION: COMPOSITION
; FILE REFERENCE: 674509-2035
; CURRENT APPLICATION NUMBER: US/09/998,284
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: PCT/IB00/00829
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: GB 9913050.2
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Chondrus crispus
US-09-998-284-1

Query Match 100.0%; Score 1644; DB 10; Length 1644;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
```

```
Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300

Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260

Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380

Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500

Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620

Qy       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
              ||||||||||||||||||||||||
Db       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
```

```
RESULT 2
US-09-824-053-30
; Sequence 30, Application US/09824053
; Patent No. US20020106725A1
;   GENERAL INFORMATION:
;        APPLICANT: Peter Stougaard
;                   Ole Cai Hansen
;        TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A
;                            METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME
;        NUMBER OF SEQUENCES: 34
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Hunton & Williams
;             STREET: 1900 K Street, N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20006-1109
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/824,053
;             FILING DATE: 03-Apr-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/669,304
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Stanislaus Aksman
;             REGISTRATION NUMBER: 28,562
;             REFERENCE/DOCKET NUMBER: <Unknown>
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (202) 955-1926
;             TELEFAX: (202) 778-2201
;             TELEX: No. US20020106725A1e
;   INFORMATION FOR SEQ ID NO: 30:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1801 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION: 84..1721
;        SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-09-824-053-30
```

```
  Query Match             77.8%;  Score 1279.2;  DB 10;  Length 1801;
  Best Local Similarity   86.1%;  Pred. No. 0;
  Matches 1416;  Conservative    0;  Mismatches  228;  Indels    0;  Gaps    0;

Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
              ||||||||| | || || || ||||| ||||| ||||| ||||| |||||||| || ||
Db         84 ATGGCTACTCTTCCTCAGAAAGACCCCGGTTATATTGTAATTGATGTCAACGCGGGCACC 143

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
               | |||||||| |||||| |  | || |||||||||||| || |||||| |  | ||||||
Db        144 GCGGACAAGCCGGACCCACGTCTCCCCTCCATGAAGCAGGGCTTCAACCGCCGCTGGATT 203

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
              || || ||||||||||||||| ||||| |||||||| ||||||||||||||||||| |
Db        204 GGAACTAATATCGATTTCGTTTATGTCGTGTACACTCCTCAAGGTGCTTGTACTGCACTT 263

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
              ||| | |||||||||||||||||||| ||||| ||||| ||||| ||||| || || ||
Db        264 GACCGTGCTATGGAAAAGTGTTCTCCCGGTACAGTCAGGATCGTCTCTGGCGGCCATTGC 323

Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
              ||||| |||||||| || |||||||| |||||||| || |||||||| |||||| | |||
Db        324 TACGAGGACTTCGTATTTGACGAATGCGTCAAGGCCATCATCAACGTCACTGGTCTCGTT 383

Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
              ||   |||||| |||||||||| |||||||||||     || || || || |||||
Db        384 GAGAGTGGTTATGACGACGATAGGGGTTACTTCGTCAGCAGTGGAGATACAAATTGGGGC 443

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
              |||||||||||||||||||||||||||||| |||||| | || || |||||||| ||||||
Db        444 TCCTTCAAGACCTTGTTCAGAGACCACGGAAGAGTTCTTCCCGGGGGTTCCTGCTACTCC 503

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              |||||  |.||||| |||||||||| |||||||||| |||||||| | ||||| ||
Db        504 GTCGGCCTCGGTGGCCACATTGTCGGCGGAGGTGACGGCATTTTGGCCCGCTTGCATGGC 563

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              | || |||||||| |   ||| || || || |||||||||||||| | |||||||||
Db        564 CTCCCCGTCGATTGGCTCAGCGGCGTGGAGGTCGTCGTTAAGCCAGTCCTCACCGAAGAC 623

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              || || || ||||| || ||||| |||||||||| |||||||| ||| | ||||||||
Db        624 TCGGTACTCAAGTATGTGCACAAAGATTCCGAAGGCAACGACGGGGAGCTCTTTTGGGCA 683

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
              ||||| ||||| ||||| || ||||| || || ||||||||||||||||||||||||||
Db        684 CACACAGGTGGCGGTGGCGGAAACTTTGGAATCATCACCAAATACTACTTCAAGGATTTG 743

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
              || ||||||||| | || |||||||| || || |||||||||   ||||||||||||||
Db        744 CCCATGTCTCCACGGGGCGTCATCGCATCAAATTTACACTTCAGCTGGGACGGTTTCACG 803

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              |||||||||||||| |||||||||| |||||||||H  | || ||||||||||||||||
Db        804 AGAGATGCCTTGCAGGATTTGTTGACAAAGTACTTCAAACTTGCCAGATGTGATTGGAAG 863

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
              ||||| ||||| ||||| |||||||||| || ||||| |||||||||| |||||||||
Db        864 AATACGGTTGGCAAGTTTCAAATCTTCCATCAGGCAGCGGAAGAGTTTGTCATGTACTTG 923

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900
              |||||||||||||| |||||||||||| | |||||||||||||| | ||||||||||||
Db        924 TATACATCCTACTCGAACGACGCCGAGCGCGAAGTTGCCCAAGACCGTCACTATCATTTG 983

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960
              |||||||||| |||||||||||||||||||||||| |||||||| ||||| || |||
Db        984 GAGGCTGACATAGAACAGATCTACAAAACATGCGAGCCCACCAAAGCGCTTGGCGGGCAT 1043

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020
              ||||| ||||| || ||||| ||  | ||  | ||||| |||||||||||||| || |||
Db       1044 GCTGGGTGGGCGCCGTTCCCCGTGCGGCCGCGCAAGAGGCACACATCCAAGACGTCGTAT 1103

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080
              |||||||||||||| |||||||||| |||||||| | |||||||| |||||||| |||
Db       1104 ATGCATGACGAGACGATGGACTACCCCTTCTACGCGCTCACTGAGACGATCAACGGCTCC 1163

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140
              || || |||||| | || ||||||||||||||| |||||||||||||| || || || |||
Db       1164 GGGCCGAATCAGCGCGGCAAGTACAAGTCTGCGTACATGATCAAGGATTTCCCGGATTTC 1223

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200
              ||||| || || |||||||||||||||||| ||||| || ||||| ||||||||||||||
Db       1224 CAGATCGACGTGATCTGGAAATACCTTACGGAGGTCCCGGACGGCTTGACTAGTGCCGAA 1283

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
              |||||||||||  | || ||||| || ||||| |||||||||||||||||||| || |||
Db       1284 ATGAAGGATGCCTTACTCCAGGTGGACATGTTTGGTGGTGAGATTCACAAGGTGGTCTGG 1343

Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320
              |||||| || ||||| || ||| | |||||||||||||||||||||||||||||||||||
Db       1344 GATGCGACGGCAGTCGCGCAGCGCGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1403

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380
              ||||||||||||||||||||||| ||| | |||||||||||||||||||||||||||||
Db       1404 CAGGAAGAAGACAAGGATGCAGTGAACCTCAAGTGGATTAGAGACTTTTACGAGGAGATG 1463

Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440
              |||||||| ||||| || |||||||||| ||||| ||||| ||||||||||||||||||
Db       1464 TATGAGCCGTATGGCGGGGTTCCAGACCCCAACACGCAGGTGGAGAGTGGTAAAGGTGTG 1523

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500
              |||||||||||||||||||||||||||||| ||||| |||||||||||||||||||| |||
Db       1524 TTTGAGGGATGCTACTTCAACTACCCGGATGTGGACTTGAACAACTGGAAGAACGGCAAG 1583

Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560
              ||||||||| | |||||||||||||||||||||||| |  | ||||||||||||||||
Db       1584 TATGGTGCCCTCGAACTTTACTTTTTGGGTAACCTGAACCGCCTCATCAAGGCCAAATGG 1643

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620
              |||||||||||  |||||||||||||||||||||||||   |||||||||||||||||||
Db       1644 TTGTGGGATCCCAACGAGATCTTCACAAACAAACAGAGCATCCCTACTAAACCTCTTAAG 1703

Qy       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
              ||||| |||||||| |||||||||
Db       1704 GAGCCCAAGCAGACGAAATAGTAG 1727
```

RESULT 3
US-09-770-149-554/c
; Sequence 554, Application US/09770149
; Patent No. US20020059663A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2024 (PARA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149

```
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER:  60/178,506
;  PRIOR FILING DATE:  2000-01-27
;  NUMBER OF SEQ ID NOS:  999
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 554
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(647)
;   OTHER INFORMATION: n = A,T,C or G
US-09-770-149-554

  Query Match             2.6%;  Score 42.6;  DB 10;  Length 647;
  Best Local Similarity   62.9%;  Pred. No. 0.0047;
  Matches   66;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1519 TACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAG 1578
              || || ||||| ||  |||| |||||||||  | |  |||     ||  |||||| |  |
Db        253 TATTTCTTGGGGAATTTGAAGAGATTGATGGATGTTAAAGCTAAGTATGATCCTGATAAT 194

Qy       1579 ATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAGGAG 1623
              ||||||  ||| | |||  ||| ||| |    | | | ||| ||
Db        193 TTCTTCAAGAACGAGCAGAGTATTCCTCCGGTTCGTGTAAAGTAG 149


RESULT 4
US-09-938-842A-1942
; Sequence 1942, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1942
;   LENGTH: 1584
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-1942

  Query Match             2.5%;  Score 41.8;  DB 9;  Length 1584;
  Best Local Similarity   50.2%;  Pred. No. 0.016;
  Matches  103;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy        315 CGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGTTCCTTCAAGACCTT 374
              |||    |  | ||   | |  |||   | |  |  |     |  || ||  |  |.||
Db        417 CGAGAGAACGGCTTGGGTTGATTCTGGTGCTACCCTCGGAGAGCTTTACTATAGAATCTC 476

Qy        375 GTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGG 434
                  |    | | | |   |  || || |  |||| | |  | |  | ||  | |||||
Db        477 TGAGAAGAGCAATGTTCTTGGATTTCCGGCGGGTTTGTCTACCACATTGGGCGTTGGTGG 536

Qy        435 TCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGATTG 494
              |||  ||  ||| |||||  ||||||  | ||   |||  | | ||||||  | ||  |||
Db        537 ACACTTTAGCGGCGGAGGATACGGTAATCTGATGAGAAAGTATGGTTTGTCGGTGGATAA 596

Qy        495 GTTATCCGGTGTTGAAGTTGTCGTT 519
                | | |||    |   | || | |
Db        597 CGTTTTCGGCTCCGGGATCGTTGAT 621


RESULT 5
US-10-170-656-3
; Sequence 3, Application US/10170656
; Publication No. US20030028927A1
; GENERAL INFORMATION:
;  APPLICANT: Lindquist, Susan K.
;  APPLICANT:  Queitsch, Christine
;  APPLICANT:  Sangster, Todd A.
;  TITLE OF INVENTION: Methods and Compositions for Revealing Hidden Generic Variation ir
;  FILE REFERENCE: ARCD:377US1
;  CURRENT APPLICATION NUMBER: US/10/170,656
;  CURRENT FILING DATE:  2002-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,211
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/379,484
;  PRIOR FILING DATE: 2002-05-10
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 2397
;   TYPE: DNA
;   ORGANISM: Wheat
US-10-170-656-3

  Query Match             2.5%;  Score 41.8;  DB 9;  Length 2397;
  Best Local Similarity   52.6%;  Pred. No. 0.022;
  Matches   91;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy       1409 CTAACACTCAGGTTGAGAGTGGTAAAGGTGTTTTTGAGGGATGCTACTTCAACTACCCTG 1468
              || ||  | | | ||||  ||  ||    | |  |||||   || | ||  |  |   ||
Db        768 CTGACGATGAAGATGAGGATGAGAAGAAGGATACTGAGGAGGGCAAGTTTGAGGAAATTG 827

Qy       1469 ATGTTGACTTGAACAACTGGAAGAACGGTAAGTATGGTGCCTTGGAACTTTACTTTTTGG 1528
              |||  ||   | |  |   | ||||    ||| |      |  |||  |||       |
Db        828 ATGAAGAGAAGGAAGAAAAGGAGAAGAAAAAGAAGAAGATCAAGGAGGTTTCTCACGAGT 887

Qy       1529 GTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAGATC 1581
              | ||||||| ||    |   ||| |||  ||| || || | ||| | ||||||
Db        888 GGAACCTGATCAACAAGCAGAAGCCCATCTGGATGAGGAAGCCTGAGGAGATC 940


RESULT 6
US-09-938-842A-2266
; Sequence 2266, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2266
;   LENGTH: 1620
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-2266

  Query Match             2.5%;  Score 41.4;  DB 9;  Length 1620;
  Best Local Similarity   53.4%;  Pred. No. 0.022;
```

```
  Matches   87;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        357 GGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTA 416
              || ||  | ||  |  |.|   | |  | | |     |  || || |  ||    |||
Db        486 GGTTTATTACAGAATTTGGGAGAAAAGCAAAGTCCATGGATTTCCCGCCGGAGTTTGTCC 545

Qy        417 CTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCA 476
              | | || |,|| | ||||| || ||   |||||| ||| |||||| | ||   ||
Db        546 CACGGTTGGTGTTGGTGGACATATAAGCGGTGGTGGTTACGGTAATATGTTGAGGAAATT 605

Qy        477 CGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTT 519
              ||| ||  |||| |||   ||   || ||   |  | || | |
Db        606 CGGATTATCAGTTGATAACTTGATCGATGCGAAGATCGTTGAT 648


RESULT 7
US-09-938-842A-281
; Sequence 281, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 281
;   LENGTH: 1599
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-281

  Query Match             2.4%;  Score 40.2;  DB 9;  Length 1599;
  Best Local Similarity   57.6%;  Pred. No. 0.055;
  Matches   72;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        393 AGTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGG 452
              ||  || || |  ||    ||   | |||||||  | || || ||  ||  ||| |||||
Db        516 AGCGTTTCCCGCCGGAGTTTGCCCCACCGTCGGCGTCGGAGGACATTTTAGCGGCGGAGG 575

Qy        453 TGACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGT 512
                |||||| |||    |||   ||||||||  |||  |||    |    | |      ||
Db        576 ATACGGTACTTTACTGAGAAAACACGGTTTAGCAGCTGATCACGTGATAGACGCACGTGT 635

Qy        513 TGTCG 517
              |||||
Db        636 TGTCG 640


RESULT 8
US-09-258-031B-70
; Sequence 70, Application US/09258031B
; Patent No. US20020168735A1
;  GENERAL INFORMATION:
;    APPLICANT:  STUIVER, Maarten Hendrik
;    APPLICANT:  CUSTERS, Jerome Humbertina Henricus Victor
;    APPLICANT:  SELA-BURLAGE, Marianne Beatrix
;    APPLICANT:  MELCHERS, Leo Sjoerd
;    APPLICANT:  VAN DEVENTER-TROOST, Johanna Pieternella
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    TITLE OF INVENTION:  ANTIFUNGAL PROTEINS, DNA CODING
;    TITLE OF INVENTION:  THEREFOR, AND HOSTS INCORPORATING
;    TITLE OF INVENTION:  SAME.
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LADAS & PARRY
;      STREET:  26 WEST 61 STREET
;      CITY:  NEW YORK
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10023 - 7604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.25" Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS 95
;      SOFTWARE:  WORDPERFECT 8
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/258,031B
;      FILING DATE:  25-FEB-1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP97/04923
;      FILING DATE:  04-SEP-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP97200831.2
;      FILING DATE:  19-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP96202466.7
;      FILING DATE:  04-SEP-1996
;  INFORMATION FOR SEQ ID NO:  70:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1757 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Arabidopsis thaliana
;      STRAIN:  Colombia
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..570, 801..1754)
US-09-258-031B-70

  Query Match             2.4%;  Score 39.2;  DB 9;  Length 1757;
  Best Local Similarity   53.2%;  Pred. No. 0.13;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        393 AGTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGG 452
              || ||  || |  ||     |||  | | || ||| | ||||| || ||     |||||||
Db        438 AGCTTATCCCGCCGGAATTTGTCCCACGGTTGGTGTCGGTGGCCATATCAGTGGTGGAGG 497

Qy        453 TGACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGT 512
              | ||||||   ||   |||    ||||| |  | || |||         || ||    | |
Db        498 TTACGGTAACATGATGAGAAAATACGGTCTCACCGTAGATAATACCATCGATGCAAGAAT 557

Qy        513 TGTCGTTAAGCCAGTCTTGACCGAAGACTCTGTTCT 548
              ||||        |   | |  ||   |||| || |
Db        558 GGTCGACGTTAATGGTATAATTGATATCTCTATTTT 593


RESULT 9
US-09-258-031B-72
; Sequence 72, Application US/09258031B
; Patent No. US20020168735A1
;  GENERAL INFORMATION:
;    APPLICANT:  STUIVER, Maarten Hendrik
;    APPLICANT:  CUSTERS, Jerome Humbertina Henricus Victor
;    APPLICANT:  SELA-BURLAGE, Marianne Beatrix
;    APPLICANT:  MELCHERS, Leo Sjoerd
;    APPLICANT:  VAN DEVENTER-TROOST, Johanna Pieternella
```

; APPLICANT: LAGEWEG, Wessel
; APPLICANT: PONSTEIN, Anne Silene
; APPLICANT: LAGEWEG, Wessel
; APPLICANT: PONSTEIN, Anne Silene
; TITLE OF INVENTION: ANTIFUNGAL PROTEINS, DNA CODING
; TITLE OF INVENTION: THEREFOR, AND HOSTS INCORPORATING
; TITLE OF INVENTION: SAME.
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LADAS & PARRY
; STREET: 26 WEST 61 STREET
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10023 - 7604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.25" Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: WORDPERFECT 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/258,031B
; FILING DATE: 25-FEB-1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP97/04923
; FILING DATE: 04-SEP-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP97200831.2
; FILING DATE: 19-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP96202466.7
; FILING DATE: 04-SEP-1996
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1527 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis thaliana
; STRAIN: Colombia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1524
US-09-258-031B-72

Query Match 2.3%; Score 38.6; DB 9; Length 1527;
Best Local Similarity 56.8%; Pred. No. 0.18;
Matches 71; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

```
Qy        393 AGTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGG 452
              || ||  || |  ||    |||  | | || ||| | ||||| || ||    ||||||||
Db        438 AGCTTATCCCGCCGGAATTTGTCCCACGGTTGGTGTCGGTGGCCATATCAGTGGTGGAGG 497

Qy        453 TGACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGT 512
              | ||||||   ||   |||    ||||| |  | || |||         || ||    | |
Db        498 TTACGGTAACATGATGAGAAAATACGGTCTCACCGTAGATAATACCATCGATGCAAGAAT 557

Qy        513 TGTCG 517
               ||||
Db        558 GGTCG 562
```

RESULT 10
US-09-938-842A-1765
; Sequence 1765, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1765
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1765

Query Match 2.3%; Score 38.6; DB 9; Length 1584;
Best Local Similarity 56.8%; Pred. No. 0.18;
Matches 71; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

```
Qy        393 AGTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGG 452
              || ||  || |  ||    |||  | | || ||| | ||||| || ||    ||||||||
Db        495 AGCTTATCCCGCCGGAATTTGTCCCACGGTTGGTGTCGGTGGCCATATCAGTGGTGGAGG 554

Qy        453 TGACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGT 512
              | ||||||   ||   |||    ||||| |  | || |||         || ||    | |
Db        555 TTACGGTAACATGATGAGAAAATACGGTCTCACCGTAGATAATACCATCGATGCAAGAAT 614

Qy        513 TGTCG 517
               ||||
Db        615 GGTCG 619
```

RESULT 11
US-09-938-842A-1762
; Sequence 1762, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1762
; LENGTH: 1599
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1762

Query Match 2.3%; Score 37.2; DB 9; Length 1599;
Best Local Similarity 50.0%; Pred. No. 0.54;
Matches 93; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```
Qy        297 GGTTGAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTG 356
              ||||||   || ||  |||    | || |  |          ||   | |  || |
Db        420 GGTTGACGTTGATTTGGACAGTAACAGTGCGTGGGCTCATGCTGGTGCTACCATCGGAGA 479
```

```
Qy        357 GGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTA 416
              || ||  | || || |      | |  |||       | ||| || |  ||||  || |
Db        480 GGTTTATTACAGGATCCAAGAGAAAAGCCAAACCCATGGTTTTCCGGCCGGTTTATGCTC 539

Qy        417 CTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCA 476
                | | ||  | ||||| ||| | ||||||||||   |||||    ||   ||   |
Db        540 AAGCCTTGGCATCGGTGGCCACTTAGTCGGTGGAGCGTACGGTTCCATGATGAGGAAGTT 599

Qy        477 CGGTTT 482
              ||||
Db        600 CGGTCT 605
```

RESULT 12
US-09-801-368-187/c
; Sequence 187, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 187
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-187

Query Match 2.2%; Score 36.4; DB 10; Length 861;
Best Local Similarity 54.5%; Pred. No. 0.64;
Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
Qy       1143 GATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAAAT 1202
              |||||  ||        |  || |||| ||| |||    || | | ||||   ||  |   |
Db        668 GATTGTTGCGGATGTGGTATTCCTTGCTGTGGTCGAGGATGCTGTGACATTTGTTGCTGT 609

Qy       1203 GAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGGGA 1262
              ||    ||||  | |    ||||  |    | | || || | ||   ||   ||  || ||
Db        608 GACATTTGCTGCTGTGGTTGTTGTTGCTGTTGCTGTTGTGGTTGTGAATTCTGCGTGTGA 549

Qy       1263 TGCTACTGCAGTTG 1276
              || | |||| ||||
Db        548 TGTTGCTGCTGTTG 535
```

RESULT 13
US-09-258-031B-19
; Sequence 19, Application US/09258031B
; Patent No. US20020168735A1
; GENERAL INFORMATION:
; APPLICANT: STUIVER, Maarten Hendrik
; APPLICANT: CUSTERS, Jerome Humbertina Henricus Victor
; APPLICANT: SELA-BURLAGE, Marianne Beatrix
; APPLICANT: MELCHERS, Leo Sjoerd
; APPLICANT: VAN DEVENTER-TROOST, Johanna Pieternella
; APPLICANT: LAGEWEG, Wessel
; APPLICANT: PONSTEIN, Anne Silene
; APPLICANT: LAGEWEG, Wessel
; APPLICANT: PONSTEIN, Anne Silene
; TITLE OF INVENTION: ANTIFUNGAL PROTEINS, DNA CODING
; TITLE OF INVENTION: THEREFOR, AND HOSTS INCORPORATING
; TITLE OF INVENTION: SAME.
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LADAS & PARRY
; STREET: 26 WEST 61 STREET
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10023 - 7604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.25" Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: WORDPERFECT 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/258,031B
; FILING DATE: 25-FEB-1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP97/04923
; FILING DATE: 04-SEP-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP97200831.2
; FILING DATE: 19-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP96202466.7
; FILING DATE: 04-SEP-1996
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1589 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helianthus annuus
; STRAIN: Zebulon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1590
US-09-258-031B-19

Query Match 2.2%; Score 36.4; DB 9; Length 1589;
Best Local Similarity 55.6%; Pred. No. 0.98;
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
Qy        394 GTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGGT 453
              | ||| || |  ||    |||    | || ||  | ||||| ||  ||   ||||| ||
Db        496 GGTTTTCCGGCAGGGGTTTGTCCAACGGTTGGCGTTGGTGGGCATTTTAGTGGTGGTGGG 555

Qy        454 GACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGTT 513
              | |||| ||||   |||    | |||||| | || |||    |    | || | || |
Db        556 TATGGTAATTTGATGAGAAAATATGGTTTGTCGGTTGATAATATTGTTGATGCTCAAATA 615

Qy        514 GTCGTT 519
              | | |
Db        616 ATAGAT 621
```

RESULT 14
US-09-258-031B-15
; Sequence 15, Application US/09258031B
; Patent No. US20020168735A1
; GENERAL INFORMATION:

```
;    APPLICANT:  STUIVER, Maarten Hendrik
;    APPLICANT:  CUSTERS, Jerome Humbertina Henricus Victor
;    APPLICANT:  SELA-BURLAGE, Marianne Beatrix
;    APPLICANT:  MELCHERS, Leo Sjoerd
;    APPLICANT:  VAN DEVENTER-TROOST, Johanna Pieternella
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    TITLE OF INVENTION:  ANTIFUNGAL PROTEINS, DNA CODING
;    TITLE OF INVENTION:  THEREFOR, AND HOSTS INCORPORATING
;    TITLE OF INVENTION:  SAME.
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LADAS & PARRY
;      STREET:  26 WEST 61 STREET
;      CITY:  NEW YORK
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10023 - 7604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.25" Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS 95
;      SOFTWARE:  WORDPERFECT 8
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/258,031B
;      FILING DATE:  25-FEB-1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP97/04923
;      FILING DATE:  04-SEP-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP97200831.2
;      FILING DATE:  19-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP96202466.7
;      FILING DATE:  04-SEP-1996
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1784 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Helianthus annuus
;      STRAIN:  Zebulon
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  21..1608
US-09-258-031B-15

  Query Match             2.2%;  Score 36.4;  DB 9;  Length 1784;
  Best Local Similarity   55.6%;  Pred. No. 1.1;
  Matches   70;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        394 GTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGGT 453
              | ||| || |  ||    |||    | || ||  | ||||| ||  ||   ||||| ||
Db        516 GGTTTTCCGGCAGGGGTTTGTCCAACGGTTGGCGTTGGTGGGCATTTTAGTGGTGGTGGG 575

Qy        454 GACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGTT 513
               | |||| ||||   |||    | |||||| | || |||    |    | || | || |
Db        576 TATGGTAATTTGATGAGAAAATATGGTTTGTCGGTTGATAATATTGTTGATGCTCAAATA 635

Qy        514 GTCGTT 519
               | | |
Db        636 ATAGAT 641


RESULT 15
```

```
US-10-123-155-464
; Sequence 464, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C30
;  CURRENT APPLICATION NUMBER: US/10/123,155
;  CURRENT FILING DATE:  2002-04-15
; Prior Application removed -  See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 464
;   LENGTH: 941
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-123-155-464

  Query Match             2.2%;  Score 35.8;  DB 9;  Length 941;
  Best Local Similarity   8.4%;  Pred. No. 1.1;
  Matches   45;  Conservative  156;  Mismatches  336;  Indels    0;  Gaps    0;

Qy         18 AAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACTCCAGACAAGCCTGACCC 77
              :|:  |:   |    : :: : :      : :  :: ::|:    :|::  | | :|: :
Db        351 VAHELAHQWFGNLVTMEWWNDLWLNEGFAKFMEFVSVSVTHPELKVGDYFFGKCFDAMEV 410

Qy         78 AAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATTGGTACCAACATCGATTT 137
              :|      : | :   |    : ::  ::  |    : :      :  : : :   : :
Db        411 DALNSSHPVSTPVENPAQIREMFDDVSYDKGACILNMLREYLSADAFKSGIVQYLQKHSY 470

Qy        138 CGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTGGACAGAGCTATGGAAAA 197
              |:    : :    |  : :  ::|  :      : :::: |: ::  :  :   :
Db        471 KNTKNEDLWDSMASICPTDGVKGMDGFCSRSQHSSSSSHWHQEGVDVKTMMNTWTLQRGF 530

Qy        198 GTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGTTACGAAGACTTCGTTTT 257
              |     : :::   ::::  :   :| :  ::     |::  ::::   :|  :
Db        531 PLITITVRGRNVHMKQEHYMKGSDGAPDTGYLWHVPLTFITSKSNMVHRFLLKTKTDVLI 590

Qy        258 CGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGA 317
               : : :    : | : ::: :  : : |   :||:| :   :: :          ::
Db        591 LPEEVEWIKFNVGMNGYYIVHYEDDGWDSLTGLLKGTHTAVSSNDRASLINNAFQLVSIG 650

Qy        318 CGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGTTCCTTCAAGACCTTGTT 377
               :| :   : :: | :     |      ::  : :::   : |   |       ::
Db        651 KLSIEKALDLSLYLKHETEIMPVFQGLNELIPMYKLMEKRDMNEVETQFKAFLIRLLRDL 710

Qy        378 CAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCA 437
               ::  :   ::  :         | ::   |: :   :   : :   :   ::| |
Db        711 IDKQTWTDEGSVSEQMLRSELLLLACVHNYQPCVQRAEGYFRKWKESNGNLSLPVDVTLA 770

Qy        438 CATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTT 497
               :   | :  | :  :: :      | ::    |  :   : :     :  : :: :|
Db        771 VFAVGAQSTEGWDFLYSKYQFSLSSTEKSQIEFALCRTQNKEKLQWLLDESFKGDKIKTQ 830

Qy        498 ATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGACTCTGTTCTTAAGTA 554
                  |   : : :       ::   : : :    :   |: : |||     :|:
Db        831 EFPQILTLIGRNPVGYPLAWQFLRKNWNKLVQKFELGSSSIAHMVMGTTNQFSTRTR 887
```

Search completed: June 15, 2003, 06:21:33
Job time : 260 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 01:27:17 ; Search time 76 Seconds
(without alignments)
6633.903 Million cell updates/sec

Title: US-09-998-284-1
Perfect score: 1644
Sequence: 1 atggctactttgccacaaaa.....ctaagcagactaaatagtag 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1279.2	77.8	1801	4	US-08-669-304-30	Sequence 30, Appl
c	2	60	3.6	80161	3	US-09-036-987A-1	Sequence 1, Appli
c	3	60	3.6	80161	4	US-09-370-700-1	Sequence 1, Appli
	4	36.4	2.2	3680	4	US-09-647-390-15	Sequence 15, Appl
	5	35.8	2.2	7766	4	US-09-125-619-3	Sequence 3, Appli
c	6	35	2.1	1632	1	US-08-155-906-1	Sequence 1, Appli
	7	33.8	2.1	1981	4	US-09-647-390-17	Sequence 17, Appl
c	8	33.4	2.0	636	1	US-07-836-642-1	Sequence 1, Appli
c	9	33.4	2.0	636	1	US-08-169-563-1	Sequence 1, Appli
c	10	33.4	2.0	636	1	US-08-403-379A-5	Sequence 5, Appli
c	11	33.4	2.0	636	2	US-08-929-414-5	Sequence 5, Appli
	12	33.4	2.0	1553	4	US-09-217-490-1	Sequence 1, Appli
	13	33.2	2.0	1458	4	US-09-480-921B-3	Sequence 3, Appli
c	14	33.2	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
c	15	33	2.0	490	2	US-08-630-822A-92	Sequence 92, Appl
c	16	33	2.0	490	2	US-09-005-069-92	Sequence 92, Appl
c	17	33	2.0	490	4	US-09-171-156A-41	Sequence 41, Appl
	18	32.6	2.0	1800	3	US-09-039-773A-1	Sequence 1, Appli
c	19	32.2	2.0	6585	3	US-08-746-111-4	Sequence 4, Appli
	20	32	1.9	1749	4	US-08-481-190-17	Sequence 17, Appl
	21	32	1.9	1749	5	PCT-US93-00869-17	Sequence 17, Appl
	22	32	1.9	1764	4	US-08-481-190-14	Sequence 14, Appl
	23	32	1.9	1764	5	PCT-US93-00869-14	Sequence 14, Appl
	24	32	1.9	2349	2	US-08-974-546-2	Sequence 2, Appli
	25	32	1.9	4597	4	US-09-221-017B-1093	Sequence 1093, Ap
	26	31.8	1.9	1206	1	US-08-318-947A-1	Sequence 1, Appli
	27	31.8	1.9	1206	2	US-08-795-303-1	Sequence 1, Appli
	28	31.8	1.9	1303	1	US-08-844-154-1	Sequence 1, Appli
	29	31.8	1.9	1303	3	US-09-126-192A-1	Sequence 1, Appli
	30	31.8	1.9	1344	2	US-08-902-585-1	Sequence 1, Appli
	31	31.8	1.9	6156	4	US-08-891-640-1	Sequence 1, Appli
	32	31.8	1.9	6156	4	US-09-723-535-3	Sequence 3, Appli
	33	31.8	1.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	34	31.8	1.9	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	35	31.4	1.9	703	4	US-08-858-207A-21	Sequence 21, Appl
	36	31.4	1.9	1546	3	US-08-961-083-21	Sequence 21, Appl
	37	31.4	1.9	6854	4	US-08-961-527-200	Sequence 200, App
c	38	31.4	1.9	7577	4	US-08-961-527-46	Sequence 46, Appl
c	39	31.4	1.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	40	31.4	1.9	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	41	31.2	1.9	915	5	PCT-US94-01149-56	Sequence 56, Appl
	42	31.2	1.9	2424	4	US-09-234-393-1	Sequence 1, Appli
	43	31.2	1.9	2424	4	US-09-360-545-15	Sequence 15, Appl
	44	31.2	1.9	2424	4	US-09-865-171-1	Sequence 1, Appli
	45	31.2	1.9	2424	4	US-09-398-395A-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-08-669-304-30
; Sequence 30, Application US/08669304
; Patent No. 6251626
; GENERAL INFORMATION:
; APPLICANT: Peter Stougaard
; APPLICANT: Ole Cai Hansen
; TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A
; TITLE OF INVENTION: METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hunton & Williams
; STREET: 1900 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006-1109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,304
; FILING DATE: 12 July 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,910
; FILING DATE: 7 June 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stanislaus Aksman
; REGISTRATION NUMBER: 28,562
; REFERENCE/DOCKET NUMBER: 320.000003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 955-1926
; TELEFAX: (202) 778-2201
; TELEX: No. 6251626e
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1801 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 84..1721
US-08-669-304-30

Query Match 77.8%; Score 1279.2; DB 4; Length 1801;

Best Local Similarity 86.1%; Pred. No. 0;
Matches 1416; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

```
Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
              ||||||||| | || || || ||||| ||||| ||||| ||||| ||||||||| || ||
Db         84 ATGGCTACTCTTCCTCAGAAAGACCCCGGTTATATTGTAATTGATGTCAACGCGGGCACC 143

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
               | |||||||| |||||| |   | || |||||||||| || |||||| |   | ||||||
Db        144 GCGGACAAGCCGGACCCACGTCTCCCCTCCATGAAGCAGGGCTTCAACCGCCGCTGGATT 203

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
              || || || |||||||||||||| ||||| |||||||| |||||||||||||||||   |
Db        204 GGAACTAATATCGATTTCGTTTATGTCGTGTACACTCCTCAAGGTGCTTGTACTGCACTT 263

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
              ||| | |||||||||||||||||||| ||||| ||||| ||||| ||||| || || ||
Db        264 GACCGTGCTATGGAAAAGTGTTCTCCCGGTACAGTCAGGATCGTCTCTGGCGGCCATTGC 323

Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
              ||||| |||||||| || |||||||| |||||||| || |||||||| |||||| | |||
Db        324 TACGAGGACTTCGTATTTGACGAATGCGTCAAGGCCATCATCAACGTCACTGGTCTCGTT 383

Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
              ||    |||||| |||||||||| ||||||||||       || || || || |||||
Db        384 GAGAGTGGTTATGACGACGATAGGGGTTACTTCGTCAGCAGTGGAGATACAAATTGGGGC 443

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
              |||||||||||||||||||||||||||| |||||| | || || |||||||| ||||||
Db        444 TCCTTCAAGACCTTGTTCAGAGACCACGGAAGAGTTCTTCCCGGGGGGTTCCTGCTACTCC 503

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              |||||   | ||||| |||||||||| |||||||||| ||||||||| | |||||  ||
Db        504 GTCGGCCTCGGTGGCCACATTGTCGGCGGAGGTGACGGCATTTTGGCCCGCTTGCATGGC 563

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              | || ||||||||| |   ||| || || || |||||||||||||| | |||||||||
Db        564 CTCCCCGTCGATTGGCTCAGCGGCGTGGAGGTCGTCGTTAAGCCAGTCCTCACCGAAGAC 623

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              || || || ||||| || ||||| |||||||||| |||||||| ||| | ||||||||
Db        624 TCGGTACTCAAGTATGTGCACAAAGATTCCGAAGGCAACGACGGGGAGCTCTTTTGGGCA 683

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
              ||||| ||||| ||||| || ||||| || || ||||||||||||||||||||||||||||
Db        684 CACACAGGTGGCGGTGGCGGAAACTTTGGAATCATCACCAAATACTACTTCAAGGATTTG 743

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
              || ||||||||| | || |||||||| || || |||||||||    ||||||||||||||
Db        744 CCCATGTCTCCACGGGGCGTCATCGCATCAAATTTACACTTCAGCTGGGACGGTTTCACG 803

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              |||||||||||||| |||||||||||| |||||||||||   | || ||||||||||||||
Db        804 AGAGATGCCTTGCAGGATTTGTTGACAAAGTACTTCAAACTTGCCAGATGTGATTGGAAG 863

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
              ||||| ||||| ||||| |||||||||| || ||||| |||||||||| |||||||||
Db        864 AATACGGTTGGCAAGTTTCAAATCTTCCATCAGGCAGCGGAAGAGTTTGTCATGTACTTG 923

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900
              |||||||||||||| |||||||||||| | |||||||||||||| | ||||||||||||
Db        924 TATACATCCTACTCGAACGACGCCGAGCGCGAAGTTGCCCAAGACCGTCACTATCATTTG 983

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960
              |||||||||||  |||||||||||||||||||||||||||| |||||||| |||||  ||| 
Db        984 GAGGCTGACATAGAACAGATCTACAAAACATGCGAGCCCACCAAAGCGCTTGGCGGGCAT 1043

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020
              |||||  |||||  || |||||  ||   | ||   | |||||  ||||||||||||||  || |||
Db       1044 GCTGGGTGGGCGCCGTTCCCCGTGCGGCCGCGCAAGAGGCACACATCCAAGACGTCGTAT 1103

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080
              |||||||||||||| |||||||||| ||||||||  | ||||||||  |||||||| |||
```

```
Db       1104 ATGCATGACGAGACGATGGACTACCCCTTCTACGCGCTCACTGAGACGATCAACGGCTCC 1163

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140
              || || |||||| | || ||||||||||||||| ||||||||||||||| || || |||
Db       1164 GGGCCGAATCAGCGCGGCAAGTACAAGTCTGCGTACATGATCAAGGATTTCCCGGATTTC 1223

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200
              ||||| || || |||||||||||||||| ||||| || ||||| ||||||||||||||||
Db       1224 CAGATCGACGTGATCTGGAAATACCTTACGGAGGTCCCGGACGGCTTGACTAGTGCCGAA 1283

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
              |||||||||||  | || ||||| || ||||| |||||||||||||||||||| || |||
Db       1284 ATGAAGGATGCCTTACTCCAGGTGGACATGTTTGGTGGTGAGATTCACAAGGTGGTCTGG 1343

Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320
              ||||| || ||||| || ||| | ||||||||||||||||||||||||||||||||||||
Db       1344 GATGCGACGGCAGTCGCGCAGCGCGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1403

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380
              ||||||||||||||||||||||||| ||| | ||||||||||||||||||||||||||||
Db       1404 CAGGAAGAAGACAAGGATGCAGTGAACCTCAAGTGGATTAGAGACTTTTACGAGGAGATG 1463

Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440
              |||||||| ||||| || |||||||||| ||||| ||||| ||||||||||||||||||
Db       1464 TATGAGCCGTATGGCGGGGTTCCAGACCCCAACACGCAGGTGGAGAGTGGTAAAGGTGTG 1523

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500
              |||||||||||||||||||||||||| ||||| ||||||||||||||||||||||| |||
Db       1524 TTTGAGGGATGCTACTTCAACTACCCGGATGTGGACTTGAACAACTGGAAGAACGGCAAG 1583

Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560
              ||||||||| | |||||||||||||||||||||||||| |   | ||||||||||||||
Db       1584 TATGGTGCCCTCGAACTTTACTTTTTGGGTAACCTGAACCGCCTCATCAAGGCCAAATGG 1643

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620
              |||||||||| |||||||||||||||||||||||||    ||||||||||||||||||||
Db       1644 TTGTGGGATCCCAACGAGATCTTCACAAACAAACAGAGCATCCCTACTAAACCTCTTAAG 1703

Qy       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
              ||||| |||||||| |||||||||
Db       1704 GAGCCCAAGCAGACGAAATAGTAG 1727
```

RESULT 2
US-09-036-987A-1/c
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A

; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Query Match 3.6%; Score 60; DB 3; Length 80161;
Best Local Similarity 52.3%; Pred. No. 4.5e-08;
Matches 157; Conservative 0; Mismatches 140; Indels 3; Gaps 1;

```
Qy        341 CCGGTGACACCAACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGC 400
              | || | | ||  | ||      | |   ||| ||||||   ||   ||        | |
Db      16033 CGGGCGCCGCCCTCGGGCAGGTGTACCGAACCCTGTTCAAGAACTGGGGCGTGACGATCC 15974

Qy        401 CAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTA 460
              |    ||  | |||  |   || ||    || || |||||  |||| || ||  ||||
Db      15973 CGACCGGCGCATGTCCCGGGGTGGGCGCAGGCGGGCACATCCTCGGCGGGGGATACGGCC 15914

Qy        461 TTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTA 520
              || |  |      |||||  |   |||||  |    || || || || |||||
Db      15913 CGCTGTCGCGCCGATTCGGTTCGGTCGTCGACTACCTTCAAGGCGTCGAGGTCGTCGT-- 15856

Qy        521 AGCCAGTCTTGACCGAAGACTCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACG 580
              |   | |  | |||  ||        |     | |   ||  | | |||   ||   ||
Db      15855 -GGTCGACCAGGCCGGTGAAGTGCACATCGTCGAGGCCGACCGGAACTCCACGGGCGCCG 15797

Qy        581 ACGGTGAGTTGTTTTGGGCTCACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCA 640
                || ||||  ||||| ||||| ||||| ||||| || |||||||| ||  ||||||
Db      15796 GTCACGACTTGTGGTGGGCGCACACCGGTGGCGGTGGCGGCAACTTCGGGATCGTCACCA 15737
```

RESULT 3
US-09-370-700-1/c
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 3.6%; Score 60; DB 4; Length 80161;
Best Local Similarity 52.3%; Pred. No. 4.5e-08;
Matches 157; Conservative 0; Mismatches 140; Indels 3; Gaps 1;

```
Qy        341 CCGGTGACACCAACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGC 400
              | || | | ||  | ||      | |   ||| ||||||   ||   ||        | |
Db      16033 CGGGCGCCGCCCTCGGGCAGGTGTACCGAACCCTGTTCAAGAACTGGGGCGTGACGATCC 15974

Qy        401 CAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTA 460
              |    ||  | |||  |   || ||    || || |||||  |||| || ||  ||||
Db      15973 CGACCGGCGCATGTCCCGGGGTGGGCGCAGGCGGGCACATCCTCGGCGGGGGATACGGCC 15914

Qy        461 TTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTA 520
              || |  |      |||||  |   |||||  |    || || || || |||||
Db      15913 CGCTGTCGCGCCGATTCGGTTCGGTCGTCGACTACCTTCAAGGCGTCGAGGTCGTCGT-- 15856

Qy        521 AGCCAGTCTTGACCGAAGACTCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACG 580
              |   | |  | |||  ||        |     | |   ||  | | |||   ||   ||
Db      15855 -GGTCGACCAGGCCGGTGAAGTGCACATCGTCGAGGCCGACCGGAACTCCACGGGCGCCG 15797

Qy        581 ACGGTGAGTTGTTTTGGGCTCACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCA 640
                || ||||  ||||| ||||| ||||| ||||| || |||||||| ||  ||||||
Db      15796 GTCACGACTTGTGGTGGGCGCACACCGGTGGCGGTGGCGGCAACTTCGGGATCGTCACCA 15737
```

RESULT 4
US-09-647-390-15
; Sequence 15, Application US/09647390
; Patent No. 6465636
; GENERAL INFORMATION:
; APPLICANT: Stuiver, Maarten
; APPLICANT: Custers, Jerome
; APPLICANT: Simons, Lambertus
; TITLE OF INVENTION: Pathogen-Inducible Promoter
; FILE REFERENCE: MOG 57707/UST
; CURRENT APPLICATION NUMBER: US/09/647,390
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: EP 98201024.1
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: PCT/EP99/02178
; PRIOR FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 3680
; TYPE: DNA
; ORGANISM: Helianthus annuus
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1889)
; NAME/KEY: CDS
; LOCATION: (1890)..(3503)
US-09-647-390-15

Query Match 2.2%; Score 36.4; DB 4; Length 3680;
Best Local Similarity 55.6%; Pred. No. 0.31;
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
Qy        394 GTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGGT 453
              | ||| || |  ||    |||    | || ||  | ||||| ||  ||   ||||| ||
Db       2412 GGTTTTCCGGCAGGGGTTTGTCCAACGGTTGGCGTTGGTGGGCATTTTAGTGGTGGTGGG 2471

Qy        454 GACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGTT 513
              | |||| ||||   |||    | |||||| | || |||    |    | || | || |
Db       2472 TATGGTAATTTGATGAGAAAATATGGTTTGTCGGTTGATAATATTGTTGATGCTCAAATA 2531

Qy        514 GTCGTT 519
              | | |
Db       2532 ATAGAT 2537
```

RESULT 5
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.

```
;  APPLICANT:  JING-REN, ZHANG
;  APPLICANT:  HARDHAM, JOHN M.
;  APPLICANT:  HOWELL, JERRILYN K.
;  APPLICANT:  BARBOUR, ALAN G.
;  APPLICANT:  WEINSTOCK, GEORGE M.
;  TITLE OF INVENTION:  VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
;  FILE REFERENCE:  UTSH:234
;  CURRENT APPLICATION NUMBER: US/09/125,619
;  CURRENT FILING DATE:  1999-01-27
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 7766
;   TYPE: DNA
;   ORGANISM: Borrelia burgdorferi
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (127)
;   OTHER INFORMATION: R = A OR G
US-09-125-619-3

  Query Match             2.2%;  Score 35.8;  DB 4;  Length 7766;
  Best Local Similarity   49.2%;  Pred. No. 0.78;
  Matches   94;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy       1190 CTAGTGCCGAAATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACA 1249
             ||  |||       |||||  |   |     | | | ||      || |||| || |||  | |
Db       1883 CTGCTGCTATTGGGAAGGGTAATGCGGATGATGGTGCGGATTTTGGTGATGGGATGAAGA 1942

Qy       1250 AGGTTGTTTGGGATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACC 1309
              ||| ||  |   |  |||| ||||  |||||  ||| | |       |   |      |||
Db       1943 AGGATGATCAGATTGCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGGAAAGTTTG 2002

Qy       1310 AGACATACTGGCAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTT 1369
                  |   | | ||  |||  |||| ||  |       | |||  |   ||| || || |
Db       2003 CTGTGAAGAAGGATGAGAAAGGGAAGGCTGAGGGGGCTATTAAGGGAGCTAGCGAGTTGT 2062

Qy       1370 ACGAGGAGATG 1380
                ||  || ||
Db       2063 TGGATAAGCTG 2073


RESULT 6
US-08-155-906-1/c
; Sequence 1, Application US/08155906
; Patent No. 5405777
;  GENERAL INFORMATION:
;    APPLICANT:  ICHIHARA, SHIGEYUKI
;    APPLICANT:  MIZUNO, TAKESHI
;    TITLE OF INVENTION:  ACETIC ACID ASSIMILATING GENE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/155,906
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/850,909
;      FILING DATE:  13-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5405777man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-552-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1632 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-155-906-1

  Query Match             2.1%;  Score 35;  DB 1;  Length 1632;
  Best Local Similarity   49.4%;  Pred. No. 0.51;
  Matches   86;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy       1445 AGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAGTATG 1504
             |   | | | |||   || |   |||   |  |||| |  ||    || || ||
Db       1573 AANNNTTCAGCATCACTAACTCCACCGTTATGCTTCACAAATATAAACCAGGAAAATAAT 1514

Qy       1505 GTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGGTTGT 1564
                 ||||| || | ||  || ||  |  |  |  || |||||   ||  |||||  | ||
Db       1513 TAACCTTGAAAGTCTAAGTTATGCTTTCCTGGCCCAAATTGAGATAGCGCAAATTTTGGT 1454

Qy       1565 GGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTA 1618
                | |   | |  | || || ||     |||||  | |  |     | | |  |||
Db       1453 AGAACAGTTAAAAAATGTTAACCCTGCAACAGACGAATCAACAAAGAACCGTTA 1400


RESULT 7
US-09-647-390-17
; Sequence 17, Application US/09647390
; Patent No. 6465636
; GENERAL INFORMATION:
;  APPLICANT:  Stuiver, Maarten
;  APPLICANT:  Custers, Jerome
;  APPLICANT:  Simons, Lambertus
;  TITLE OF INVENTION:  Pathogen-Inducible Promoter
;  FILE REFERENCE:  MOG 57707/UST
;  CURRENT APPLICATION NUMBER: US/09/647,390
;  CURRENT FILING DATE:  2000-09-29
;  PRIOR APPLICATION NUMBER:  EP 98201024.1
;  PRIOR FILING DATE:  1998-04-01
;  PRIOR APPLICATION NUMBER:  PCT/EP99/02178
;  PRIOR FILING DATE:  1999-03-25
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 17
;   LENGTH: 1981
;   TYPE: DNA
;   ORGANISM: Lactuca sativa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7)..(1626)
;   NAME/KEY: unsure
;   LOCATION: (372)
;   OTHER INFORMATION: replace (372, "g")
;   NAME/KEY: unsure
;   LOCATION: (379)
;   OTHER INFORMATION: replace (379, "g")
;   NAME/KEY: unsure
;   LOCATION: (786)
;   OTHER INFORMATION: replace (786, "t")
;   NAME/KEY: unsure
;   LOCATION: (1105)..(1106)
;   OTHER INFORMATION: replace (1105..1106, "ga" or "gg" or "aa")
US-09-647-390-17

  Query Match             2.1%;  Score 33.8;  DB 4;  Length 1981;
  Best Local Similarity   52.5%;  Pred. No. 1.4;
```

```
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        379 AGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCAC 438
              | |  | || ||   | ||| || | |||    ||     | || ||  | ||||| ||
Db        511 AAAAGCAACAGTCATGCTTTTCCGGCTGGCGTTTGCCCTACTGTTGGAGTTGGTGGCCAT 570

Qy        439 ATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTA 498
               ||   ||||| ||| | ||||  |||    ||    ||||  |  | || ||   |
Db        571 TTTAGTGGTGGTGGTTATGGTAACTTGATGGGAAAATACGGCCTTTCTGTTGACAATATT 630

Qy        499 TCCGGTGTTGAAGTTGTCGTT 519
              || || | |  |  ||| |
Db        631 GTCGATGCTCAGTTAATCGAT 651


RESULT 8
US-07-836-642-1/c
; Sequence 1, Application US/07836642
; Patent No. 5304371
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  Peptide for Diagnosing and Immunizing
;    TITLE OF INVENTION:  Against T. cruzi Infection
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University street
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/836,642
;      FILING DATE:  14-FEB-1992
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia A.
;      REGISTRATION NUMBER:  34,693
;      REFERENCE/DOCKET NUMBER:  REED
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206)587-0430
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Trypanosoma cruzi
;    IMMEDIATE SOURCE:
;      CLONE:  TcD
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8..628
US-07-836-642-1

  Query Match             2.0%;  Score 33.4;  DB 1;  Length 636;
  Best Local Similarity   51.0%;  Pred. No. 0.93;
  Matches   79;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        208 GGTACCGTCAGAATCGTTTCTGGTGGTCACTGTTACGAAGACTTCGTTTTCGACGAATGT 267
              ||  |||      |||  || | ||||    | | ||  |||||||  | ||  |  |
Db        555 GGCTCCGCTGATTTCGGCTCCGCTGGTTTTGGCTCCGCTGACTTCGGCTCCGCTGGTTTA 496

Qy        268 GTCAAGGCTATTATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGT 327
              | |  |||  | || ||    |||||||||      ||| || || |  || | |
Db        495 GGCCCCGCTGATTTCGACTCCGCTGGTTTGGGCTCTGCTGATTTCGGCTCCGCTGGTTTG 436

Qy        328 TACTTCGTCTCTTCCGGTGACACCAACTGGGGTTC 362
                || ||    || |||   | |    | ||| ||
Db        435 GGCTCCGCTGATTTCGGCTCCGCTGGTTTGGGCTC 401


RESULT 9
US-08-169-563-1/c
; Sequence 1, Application US/08169563
; Patent No. 5413912
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  Peptide for Diagnosing and Immunizing
;    TITLE OF INVENTION:  Against T. cruzi Infection
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University street
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple Operating System 7.1
;      SOFTWARE:  Microsoft Word For Apple, Version 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/169,563
;      FILING DATE:  17-DEC-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/836,642
;      FILING DATE:  14-FEB-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia A.
;      REGISTRATION NUMBER:  34,693
;      REFERENCE/DOCKET NUMBER:  REED-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206)587-0430
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Trypanosoma cruzi
;    IMMEDIATE SOURCE:
;      CLONE:  TcD
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8..628
US-08-169-563-1

  Query Match             2.0%;  Score 33.4;  DB 1;  Length 636;
  Best Local Similarity   51.0%;  Pred. No. 0.93;
  Matches   79;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        208 GGTACCGTCAGAATCGTTTCTGGTGGTCACTGTTACGAAGACTTCGTTTTCGACGAATGT 267
              ||  |||      |||  || | ||||    | | ||  |||||||  | ||  |  |
Db        555 GGCTCCGCTGATTTCGGCTCCGCTGGTTTTGGCTCCGCTGACTTCGGCTCCGCTGGTTTA 496

Qy        268 GTCAAGGCTATTATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGT 327
              | |  |||  | || ||    |||||||||      ||| || || |  || | |
Db        495 GGCCCCGCTGATTTCGACTCCGCTGGTTTGGGCTCTGCTGATTTCGGCTCCGCTGGTTTG 436
```

```
Qy        328 TACTTCGTCTCTTCCGGTGACACCAACTGGGGTTC 362
              || ||    || |||   | |    | ||| ||
Db        435 GGCTCCGCTGATTTCGGCTCCGCTGGTTTGGGCTC 401


RESULT 10
US-08-403-379A-5/c
; Sequence 5, Application US/08403379A
; Patent No. 5756662
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DETECTION
;    TITLE OF INVENTION:  OF T. CRUZI INFECTION
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,379A
;      FILING DATE:  14-MAR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.406
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8..628
US-08-403-379A-5

  Query Match              2.0%;  Score 33.4;  DB 1;  Length 636;
  Best Local Similarity   51.0%;  Pred. No. 0.93;
  Matches   79;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        208 GGTACCGTCAGAATCGTTTCTGGTGGTCACTGTTACGAAGACTTCGTTTTCGACGAATGT 267
              ||  |||       |||  || | ||||    | | ||  |||||||  | ||  |  |
Db        555 GGCTCCGCTGATTTCGGCTCCGCTGGTTTTGGCTCCGCTGACTTCGGCTCCGCTGGTTTA 496

Qy        268 GTCAAGGCTATTATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGT 327
              | |   |||  | || ||    |||||||||      ||| || || |  || | |
Db        495 GGCCCCGCTGATTTCGACTCCGCTGGTTTGGGCTCTGCTGATTTCGGCTCCGCTGGTTTG 436

Qy        328 TACTTCGTCTCTTCCGGTGACACCAACTGGGGTTC 362
              || ||    || |||   | |    | ||| ||
Db        435 GGCTCCGCTGATTTCGGCTCCGCTGGTTTGGGCTC 401


RESULT 11
US-08-929-414-5/c
; Sequence 5, Application US/08929414
; Patent No. 5942403
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Haughton, Raymond
;    APPLICANT:  Skeiky, Yasir A.W.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DETECTION
;    TITLE OF INVENTION:  OF T. CRUZI INFECTION
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/929,414
;      FILING DATE:  15-SEP-1997
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.406C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  636 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  8..628
US-08-929-414-5

  Query Match              2.0%;  Score 33.4;  DB 2;  Length 636;
  Best Local Similarity   51.0%;  Pred. No. 0.93;
  Matches   79;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        208 GGTACCGTCAGAATCGTTTCTGGTGGTCACTGTTACGAAGACTTCGTTTTCGACGAATGT 267
              ||  |||       |||  || | ||||    | | ||  |||||||  | ||  |  |
Db        555 GGCTCCGCTGATTTCGGCTCCGCTGGTTTTGGCTCCGCTGACTTCGGCTCCGCTGGTTTA 496

Qy        268 GTCAAGGCTATTATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGT 327
              | |   |||  | || ||    |||||||||      ||| || || |  || | |
Db        495 GGCCCCGCTGATTTCGACTCCGCTGGTTTGGGCTCTGCTGATTTCGGCTCCGCTGGTTTG 436

Qy        328 TACTTCGTCTCTTCCGGTGACACCAACTGGGGTTC 362
              || ||    || |||   | |    | ||| ||
Db        435 GGCTCCGCTGATTTCGGCTCCGCTGGTTTGGGCTC 401


RESULT 12
US-09-217-490-1
; Sequence 1, Application US/09217490
; Patent No. 6165761
; GENERAL INFORMATION:
;  APPLICANT: Schneider, Palle
;  APPLICANT: Christensen, Soren
;  APPLICANT: Dybdal, Lone
;  APPLICANT: Fuglsang, Claus Crone
;  APPLICANT: Xu, Feng
;  APPLICANT: Golightly, Elizabeth
;  TITLE OF INVENTION: Carbohydrate Oxidase And use Thereof In
;  TITLE OF INVENTION: Baking
;  FILE REFERENCE: 5421.200-US
;  CURRENT APPLICATION NUMBER: US/09/217,490
;  CURRENT FILING DATE: 1998-12-21
;  EARLIER APPLICATION NUMBER: PA 1997 01505
;  EARLIER FILING DATE: 1997-12-22
;  EARLIER APPLICATION NUMBER: PA 1998 00763
```

```
;  EARLIER FILING DATE: 1998-06-04
;  EARLIER APPLICATION NUMBER: 60/068,717
;  EARLIER FILING DATE: 1997-12-23
;  EARLIER APPLICATION NUMBER: 60/088/725
;  EARLIER FILING DATE: 1998-06-10
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1553
;   TYPE: DNA
;   ORGANISM: Microdochium nivale
;   FEATURE:
;   NAME/KEY: intron
;   LOCATION: (1012)..(1076)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1011)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1077)..(1553)
;   FEATURE:
;   NAME/KEY: mat_peptide
;   LOCATION: (67)..(1550)
US-09-217-490-1

  Query Match              2.0%;  Score 33.4;  DB 4;  Length 1553;
  Best Local Similarity   49.2%;  Pred. No. 1.7;
  Matches   88;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        341 CCGGTGACACCAACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGC 400
              |||||| |  |  | |   |  | ||    | |   |     |  ||||  | |   |
Db        389 CCGGTGCCCGCCTCGGACATCTCGCCACCGTCCTCAACGACAAGTACGGCCGTGCCATCT 448

Qy        401 CAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTA 460
              |    ||| | ||       |||||  |    || ||| ||| |   || ||   |||
Db        449 CCCACGGTACATGCCCTGGTGTCGGCATCTCCGGCCACTTTGCCCACGGCGGCTTCGGCT 508

Qy        461 TTTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTT 519
              |     |     |||||||| || | ||||| | | |   ||||||    |||||  ||
Db        509 TCAGCTCGCACATGCACGGTCTGGCTGTCGACTCGGTCGTCGGTGTCACTGTTGTTCTT 567


RESULT 13
US-09-480-921B-3
; Sequence 3, Application US/09480921B
; Patent No. 6387637
; GENERAL INFORMATION:
;  APPLICANT: Levin, Joshua Z.
;  APPLICANT:  Budziszewski, Gregory J.
;  APPLICANT:  Potter, Sharon L.
;  APPLICANT:  Wegrich, Lynette M.
;  TITLE OF INVENTION: Herbicide Target Genes and Methods
;  FILE REFERENCE: PB/5-30780A
;  CURRENT APPLICATION NUMBER: US/09/480,921B
;  CURRENT FILING DATE:  2000-01-11
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1458
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1458)
US-09-480-921B-3

  Query Match              2.0%;  Score 33.2;  DB 4;  Length 1458;
  Best Local Similarity   50.0%;  Pred. No. 1.8;
  Matches   83;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy       1342 GTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATGTATGAGCCTTATGGTGGTGTT 1401
              | | | ||| |   | |    |||  | | |  ||  || | |||  | ||||||  ||||
Db         28 GCTGATTTGGACGAGTTATCTGACAATGAAGCAGAATTGGACGAGAATGATGGTGATGTT 87

Qy       1402 CCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTTTTTGAGGGATGCTACTTCAAC 1461
              | |      |      | |||||  |  | ||  | || ||| |||  |    |||   |
Db         88 GGAAAGGAAGAAGAAGATGTTGATATGGATATGGCTGATTTAGAGACACTTAACTATGAT 147

Qy       1462 TACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAGTATGGTG 1507
              | |  |||  ||  |  ||      |||||     |  |||| ||
Db        148 GATCTCGATAATGTTTCTAAGCTGCAGAAGAGTCAGAGATATGCTG 193


RESULT 14
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match              2.0%;  Score 33.2;  DB 1;  Length 7218;
  Best Local Similarity   6.2%;  Pred. No. 5.2;
  Matches   26;  Conservative  203;  Mismatches  191;  Indels    0;  Gaps    0;

Qy       1097 GTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTCCAGATTGATGTTATCT 1156
              |||| |||   |||    |  |  | || ||  |   ||| ||    |  :: :  :
Db       1483 GTAATTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRR 1424

Qy       1157 GGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAAATGAAGGATGCTCTTC 1216
              ::::: :    :  ::::    :: ::   ::  :: :  :::: :::::: :
Db       1423 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1364

Qy       1217 TTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGGGATGCTACTGCAGTTG 1276
                :::  :: : :   :: :: ::::   : ::::  :   :::: :  :  : ::  :
```

```
Db       1363 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1304

Qy       1277 CTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGGCAGGAAGAAGACAAGG 1336
              :::::::: : :  :  :::  : :: :  ::: : :  :: :::::::::: ::::
Db       1303 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1244

Qy       1337 ATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATGTATGAGCCTTATGGTG 1396
              : : ::  ::   :::: :::  :::::     : ::::::: : : :::    : :: :
Db       1243 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1184

Qy       1397 GTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTTTTTGAGGGATGCTACT 1456
              : :    :::    :: :   :::  ::::: :: ::::: :     :::::: :  :
Db       1183 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1124

Qy       1457 TCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAGTATGGTGCCTTGGAAC 1516
                ::  :    :: : :  ::   ::: ::  :::::::: :: ::: : :: :    ::| |
Db       1123 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATC 1064
```

RESULT 15
US-08-630-822A-92/c
; Sequence 92, Application US/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,822A
; FILING DATE: 11-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 490 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-630-822A-92

Query Match 2.0%; Score 33; DB 2; Length 490;
Best Local Similarity 53.5%; Pred. No. 1.1;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy        208 GGTACCGTCAGAATCGTTTCTGGTGGTCACTGTTACGAAGACTTCGTTTTCGACGAATGT 267
              ||||  | |||    || ||  ||||| |||| | || | ||| ||||   |  ||
Db        213 GGTATAGCCAGTGGTGTATCCAGTGGTGACTGGTGCGGAAACTACGTTGGTGGTGACGTA 154

Qy        268 GTCAAGGCTATTATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGT 327
              ||||  |  |  ||   || |  ||  |    | | |   ||    || ||  | | |||
Db        153 GTCATTGGCAAGATGTCCGGTGGTGTATCCAGTAACTGGAGTGGATACAACATTGGTGGT 94
```

```
Qy        328 TACTTCGTC 336
              || | |||
Db         93 GACGTAGTC 85
```

Search completed: June 15, 2003, 02:58:25
Job time : 82 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 02:43:57 ; Search time 2460 Seconds
(without alignments)
10823.325 Million cell updates/sec

Title: US-09-998-284-1
Perfect score: 1644
Sequence: 1 atggctactttgccacaaaa.....ctaagcagactaaatagtag 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	42.8	2.6	646	17	BH695306	BH695306 BOMLJ48TR
c	2	42.8	2.6	709	17	BH667535	BH667535 BOHZJ13TR
c	3	42.8	2.6	847	17	BH432838	BH432838 BOGVZ24TF
c	4	42.6	2.6	420	17	BH813311	BH813311 SALK_0639
c	5	42.6	2.6	447	10	AV786257	AV786257 AV786257
c	6	42.6	2.6	640	10	AV784340	AV784340 AV784340
c	7	42	2.6	400	10	AV811765	AV811765 AV811765
c	8	42	2.6	440	10	AV798160	AV798160 AV798160
c	9	42	2.6	442	10	AV798159	AV798159 AV798159
	10	41.8	2.5	748	14	BQ841497	BQ841497 WHE4211_F
	11	41.4	2.5	409	14	BQ060219	BQ060219 Ts1168 Th
c	12	41.4	2.5	1101	17	CNS00LO0	AL068607 Drosophil
c	13	41	2.5	406	10	AV805562	AV805562 AV805562
	14	41	2.5	416	12	BF202701	BF202701 WHE1777_C
	15	41	2.5	569	14	BU013165	BU013165 QGJ3O15.y
	16	41	2.5	595	14	BU007777	BU007777 QGH5H07.y
	17	41	2.5	688	14	BQ990898	BQ990898 QGF21G13.
	18	41	2.5	807	17	AQ957642	AQ957642 LERAQ85TF
c	19	40.4	2.5	360	10	AV440363	AV440363 AV440363
	20	40.2	2.4	360	13	BI594749	BI594749 As_tgz_43
c	21	40.2	2.4	360	14	D76262	D76262 CELK116B6F
	22	40.2	2.4	387	12	BF428917	BF428917 WHE1710_H
	23	40.2	2.4	493	13	BM170559	BM170559 EST573082
c	24	40.2	2.4	537	13	BM277996	BM277996 As_tgz_51
	25	40.2	2.4	600	12	BF485368	BF485368 WHE2310_G
	26	40.2	2.4	633	14	BQ607036	BQ607036 BRY_2917
	27	40.2	2.4	769	10	BE414425	BE414425 SCU009.F0
c	28	40	2.4	350	13	BI594738	BI594738 As_tgz_43
c	29	40	2.4	492	9	AI995621	AI995621 701675986
c	30	40	2.4	524	13	BI594743	BI594743 As_tgz_43
c	31	40	2.4	565	12	BF049993	BF049993 As_tgz_20
	32	39.6	2.4	647	9	AL509182	AL509182 AL509182
c	33	39.6	2.4	770	17	BH478648	BH478648 BOHKR61TR
	34	39.4	2.4	360	9	AJ470390	AJ470390 AJ470390
c	35	39.4	2.4	437	17	BH611048	BH611048 SALK_0297
	36	39.4	2.4	540	9	AJ470389	AJ470389 AJ470389
	37	39.4	2.4	542	14	BQ468825	BQ468825 HM02H09r
	38	39.4	2.4	573	13	BJ471443	BJ471443 BJ471443
	39	39.4	2.4	577	14	BQ839404	BQ839404 WHE4165_G
	40	39.4	2.4	603	13	BJ479961	BJ479961 BJ479961
c	41	39.4	2.4	608	17	B05851	B05851 cSRL-72a8-u
	42	39.4	2.4	624	13	BM441131	BM441131 EBed02_SQ
c	43	39.4	2.4	632	17	BH605523	BH605523 BOHAE53TR
	44	39.4	2.4	634	10	AV938215	AV938215 AV938215
	45	39.4	2.4	649	10	AV941742	AV941742 AV941742

ALIGNMENTS

RESULT 1
BH695306/c
LOCUS BH695306 646 bp DNA linear GSS 20-FEB-2002
DEFINITION BOMLJ48TR BO_2_3_KB Brassica oleracea genomic clone BOMLJ48, DNA sequence.
ACCESSION BH695306
VERSION BH695306.1 GI:18767922
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 646)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOMLJ48TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .646

```
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOMLJ48"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      123 a    162 c    220 g    141 t
ORIGIN

  Query Match             2.6%;  Score 42.8;  DB 17;  Length 646;
  Best Local Similarity   50.5%;  Pred. No. 0.16;
  Matches  104;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy        312 CGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGTTCCTTCAAGAC 371
              || || |||    | ||   | |   |   ||    |  |  |  || || || || ||
Db        383 CGTCGGCGAGTCCGCTTGGGTGGGAGCCGGCGCCACCCTCGGCGAGGTTTACTACAGGAT 324

Qy        372 CTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGG 431
              | |   | |    |        |  || || |  ||   |||   | |||||||    ||
Db        323 ATGGGAAAAAACAAAGCTCCACGGGTTCCCCGCCGGAGTCTGCCCCACCGTCGGAGCCGG 264

Qy        432 TGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGTTTGCCAGTCGA 491
              ||||||||   ||| || ||  ||||||   ||  ||||  | ||||  |  | || ||
Db        263 AGGTCACATCAGCGGCGGCGGGTACGGTAACATGATCAGAAAGTACGGACTCTCCGTTGA 204

Qy        492 TTGGTTATCCGGTGTTGAAGTTGTCG 517
              |   |  |||  | | |  | || |
Db        203 CTACGTCACCGACGCTAAGATCGTTG 178


RESULT 2
BH667535/c
LOCUS       BH667535                709 bp    DNA     linear   GSS 19-FEB-2002
DEFINITION  BOHZJ13TR BO_2_3_KB Brassica oleracea genomic clone BOHZJ13, DNA
            sequence.
ACCESSION   BH667535
VERSION     BH667535.1  GI:18726845
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 709)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .709
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHZJ13"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      162 a    191 c    151 g    205 t
ORIGIN

  Query Match             2.6%;  Score 42.8;  DB 17;  Length 709;
  Best Local Similarity   50.0%;  Pred. No. 0.17;
  Matches  135;  Conservative    0;  Mismatches  132;  Indels    3;  Gaps    1;

Qy       1109 CTGCTTACATGATCAAGGACTTTCCAGACTTCCAGATTGATGTTATCTGGAAATACCTTA 1168
              |||   |  |  |||||  |    |||||| |   |   |    |||   |||  |   ||
Db        331 CTGGAGATTTCTTCAAGAGCAAGTCAGACTACGTCAAGAAACCAATCCCCAAAGAAGGTA 272

Qy       1169 CTGAGGTTCCTGACGGTTTGACTAGTGCCGAAATGAAGGATGCTCTTCTTCAGGTTGATA 1228
              |||   | |   |      |             | ||| | ||      | |||    |
Db        271 TGGAGAAGCTTTGGGCGACAATGTTGAAATTCAAGAACGTTGTGTGGATGCAGTGGAACC 212

Qy       1229 TGTTCGGTGGTGAGATTCACAAGGTTGTTTGGGATGCTACTGCAGTTGCTCA---GAGAG 1285
              | |||||| | |||  | | | ||     |   || || ||  || ||||   || ||
Db        211 CTTACGGTGGCGTGATGGATAGGATTCCGGCGACCGCCACGGCGTTTCCTCACCGGAAAG 152

Qy       1286 AGTACATCATCAAACTGCAGTACCAGACATACTGGCAGGAAGAAGACAAGGATGCAGTTA 1345
              || |  ||||  | ||||||   ||    |||  |||  | | |||     | || ||
Db        151 GAAACTTGTTCAAGATTCAGTACTTTACTACATGGTTGGACGCAAACACCACAGAAGCTA 92

Qy       1346 ACTTGAAGTGGATTAGAGACTTTTACGAGG 1375
              |||  ||   |||  ||  ||  ||||||||||
Db         91 GCTTAAAAATGATGAGGGAGTTTTACGAGG 62


RESULT 3
BH432838/c
LOCUS       BH432838                847 bp    DNA     linear   GSS 12-DEC-2001
DEFINITION  BOGVZ24TF BOGV Brassica oleracea genomic clone BOGVZ24, DNA
            sequence.
ACCESSION   BH432838
VERSION     BH432838.1  GI:17618559
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 847)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGVZ24TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .847
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGVZ24"
                     /clone_lib="BOGV"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      201 a    236 c    176 g    234 t
ORIGIN

  Query Match             2.6%;  Score 42.8;  DB 17;  Length 847;
  Best Local Similarity   50.0%;  Pred. No. 0.2;
  Matches  135;  Conservative    0;  Mismatches  132;  Indels    3;  Gaps    1;

Qy       1109 CTGCTTACATGATCAAGGACTTTCCAGACTTCCAGATTGATGTTATCTGGAAATACCTTA 1168
              |||   |  |  |||||  |    |||||| |   |   |    |||   |||  |   ||
Db        305 CTGGAGATTTCTTCAAGAGCAAGTCAGACTACGTCAAGAAACCAATCCCCAAAGAAGGTA 246

Qy       1169 CTGAGGTTCCTGACGGTTTGACTAGTGCCGAAATGAAGGATGCTCTTCTTCAGGTTGATA 1228
              |||   | |   |      |             | ||| | ||      | |||    |
Db        245 TGGAGAAGCTTTGGGCGACAATGTTGAAATTCAAGAACGTTGTGTGGATGCAGTGGAACC 186

Qy       1229 TGTTCGGTGGTGAGATTCACAAGGTTGTTTGGGATGCTACTGCAGTTGCTCA---GAGAG 1285
```

```
                | |||||| | |||  | | | ||     |   || || ||  || ||||   || ||
Db        185 CTTACGGTGGCGTGATGGATAGGATTCCGGCGACCGCCACGGCGTTTCCTCACCGGAAAG 126

Qy       1286 AGTACATCATCAAACTGCAGTACCAGACATACTGGCAGGAAGAAGACAAGGATGCAGTTA 1345
               || |  ||||  | ||||||  ||    |||  ||| | | |||     | || ||
Db        125 GAAACTTGTTCAAGATTCAGTACTTTACTACATGGTTGGACGCAAACACCACAGAAGCTA 66

Qy       1346 ACTTGAAGTGGATTAGAGACTTTTACGAGG 1375
               ||| ||   ||| || || ||||||||||
Db         65 GCTTAAAAATGATGAGGGAGTTTTACGAGG 36
```

RESULT 4
BH813311/c

```
LOCUS       BH813311                420 bp    DNA     linear   GSS 02-MAY-2002
DEFINITION  SALK_063965 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_063965, DNA sequence.
ACCESSION   BH813311
VERSION     BH813311.1  GI:20391766
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .420
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_063965"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT      147 a     78 c     66 g    129 t
ORIGIN
```

```
  Query Match             2.6%;  Score 42.6;  DB 17;  Length 420;
  Best Local Similarity   62.9%;  Pred. No. 0.13;
  Matches   66;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1519 TACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAG 1578
              || || ||||| ||  |||| ||||||||  | |  |||     ||  |||||| |  |
Db        408 TATTTCTTGGGGAATTTGAAGAGATTGATGGATGTTAAAGCTAAGTATGATCCTGATAAT 349

Qy       1579 ATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAGGAG 1623
              |||||| |||| |||||   || ||| |    | | | | | ||
Db        348 TTCTTCAAAAACGAACAGAGCATTCCTCCGGTTCGTGTAATGTAG 304
```

RESULT 5
AV786257/c

```
LOCUS       AV786257                447 bp    mRNA    linear   EST 28-MAR-2002
DEFINITION  AV786257 RAFL6 Arabidopsis thaliana cDNA clone RAFL06-70-H12 3',
            mRNA sequence.
ACCESSION   AV786257
VERSION     AV786257.1  GI:19805047
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). This clone is in a
            modified pBluescript vector as a SstI/XhoI insert. Please visit our
            web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="RAFL06-70-H12"
                     /clone_lib="RAFL6"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /note="Site_1: SstI; Site_2: XhoI; subjected to
                     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                     hr) treatments"
BASE COUNT      150 a     89 c     69 g    139 t
ORIGIN
```

```
  Query Match             2.6%;  Score 42.6;  DB 10;  Length 447;
  Best Local Similarity   62.9%;  Pred. No. 0.14;
  Matches   66;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1519 TACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAG 1578
              || || ||||| ||  |||| ||||||||  | |  |||     ||  |||||| |  |
Db        252 TATTTCTTGGGGAATTTGAAGAGATTGATGGATGTTAAAGCTAAGTATGATCCTGATAAT 193

Qy       1579 ATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAGGAG 1623
              |||||| |||| |||||   || ||| |    | | | | | ||
Db        192 TTCTTCAAAAACGAACAGAGCATTCCTCCGGTTCGTGTAATGTAG 148
```

RESULT 6
AV784340/c

```
LOCUS       AV784340                640 bp    mRNA    linear   EST 28-MAR-2002
DEFINITION  AV784340 RAFL5 Arabidopsis thaliana cDNA clone RAFL05-18-J21 3',
            mRNA sequence.
ACCESSION   AV784340
VERSION     AV784340.1  GI:19803130
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
```

```
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). This clone is in a
            modified pBluescript vector as a SstI/XhoI insert. Please visit our
            web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="RAFL05-18-J21"
                     /clone_lib="RAFL5"
                     /dev_stage="rosette plants"
                     /lab_host="SOLR"
                     /note="Site_1: SstI; Site_2: XhoI; subjected to
                     dehydration-treated(1,2,5,10,24 hr)"
BASE COUNT      190 a    142 c    106 g    202 t
ORIGIN

  Query Match             2.6%;  Score 42.6;  DB 10;  Length 640;
  Best Local Similarity   62.9%;  Pred. No. 0.18;
  Matches   66;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1519 TACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAG 1578
              || || ||||| ||  |||| ||||||||  | |  |||     ||  ||||||  |  |
Db        211 TATTTCTTGGGGAATTTGAAGAGATTGATGGATGTTAAAGCTAAGTATGATCCTGATAAT 152

Qy       1579 ATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAGGAG 1623
              ||||||  ||||  |||||   ||  |||  |    |  |  |  |  ||
Db        151 TTCTTCAAAAACGAACAGAGCATTCCTCCGGTTCGTGTAATGTAG 107


RESULT 7
AV811765/c
LOCUS       AV811765                400 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AV811765 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-69-I23 3',
            mRNA sequence.
ACCESSION   AV811765
VERSION     AV811765.1  GI:19845750
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="RAFL09-69-I23"
                     /clone_lib="RAFL9"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /note="Site_1: BamHI; Site_2: SalI; subjected to
                     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                     hr) treatments"
BASE COUNT      115 a     90 c     59 g    136 t
ORIGIN

  Query Match             2.6%;  Score 42;  DB 10;  Length 400;
  Best Local Similarity   66.7%;  Pred. No. 0.19;
  Matches   60;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       1519 TACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAG 1578
              || || ||||| ||  |||| ||||||||  | |  |||     | | ||||||  |  |
Db        237 TATTTCTTGGGGAATTTGAAGAGATTGATGGATGTTAAAGCTAAGAGTGATCCTGAGAAT 178

Qy       1579 ATCTTCACAAACAAACAGTCTATCCCTACT 1608
              ||||||  ||||  |||||  |||  ||  ||
Db        177 TTCTTCAAAAACGAACAGAGTATTCCCCCT 148


RESULT 8
AV798160/c
LOCUS       AV798160                440 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AV798160 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-15-D04 3',
            mRNA sequence.
ACCESSION   AV798160
VERSION     AV798160.1  GI:19832143
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .440
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="RAFL09-15-D04"
                     /clone_lib="RAFL9"
                     /dev_stage="plants at various developmental stages from
```

```
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /note="Site_1: BamHI; Site_2: SalI; subjected to
                     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                     hr) treatments"
BASE COUNT      130 a     96 c     68 g    146 t
ORIGIN

  Query Match             2.6%;  Score 42;  DB 10;  Length 440;
  Best Local Similarity   66.7%;  Pred. No. 0.21;
  Matches   60;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       1519 TACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAG 1578
              || || ||||| ||  |||| ||||||||  | |  |||      | | ||||||  |  |
Db        267 TATTTCTTGGGGAATTTGAAGAGATTGATGGATGTTAAAGCTAAGAGTGATCCTGAGAAT 208

Qy       1579 ATCTTCACAAACAAACAGTCTATCCCTACT 1608
              ||||||  ||||  |||||  |||  ||  ||
Db        207 TTCTTCAAAAACGAACAGAGTATTCCCCCT 178


RESULT 9
AV798159/c
LOCUS       AV798159                442 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AV798159 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-15-D03 3',
            mRNA sequence.
ACCESSION   AV798159
VERSION     AV798159.1  GI:19832142
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .442
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="RAFL09-15-D03"
                     /clone_lib="RAFL9"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /note="Site_1: BamHI; Site_2: SalI; subjected to
                     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                     hr) treatments"
BASE COUNT      129 a     98 c     68 g    147 t
ORIGIN

  Query Match             2.6%;  Score 42;  DB 10;  Length 442;
  Best Local Similarity   66.7%;  Pred. No. 0.21;
  Matches   60;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       1519 TACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAG 1578
              || || ||||| ||  |||| ||||||||  | |  |||      | | ||||||  |  |
Db        267 TATTTCTTGGGGAATTTGAAGAGATTGATGGATGTTAAAGCTAAGAGTGATCCTGAGAAT 208

Qy       1579 ATCTTCACAAACAAACAGTCTATCCCTACT 1608
              ||||||  ||||  |||||  |||  ||  ||
Db        207 TTCTTCAAAAACGAACAGAGTATTCCCCCT 178


RESULT 10
BQ841497
LOCUS       BQ841497                748 bp    mRNA    linear   EST 12-AUG-2002
DEFINITION  WHE4211_F11_L20ZS Aegilops speltoides pre-meiotic anther cDNA
            library Aegilops speltoides cDNA clone WHE4211_F11_L20, mRNA
            sequence.
ACCESSION   BQ841497
VERSION     BQ841497.1  GI:22210906
KEYWORDS    EST.
SOURCE      Aegilops speltoides.
  ORGANISM  Aegilops speltoides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Aegilops.
REFERENCE   1  (bases 1 to 748)
  AUTHORS   Akhunov,E., Anderson,O.D., Chao,S., Crossman,C., Devorak,J., Lazo
            ,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Pre-meiotic anther cDNA library from Aegilops speltoides
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES             Location/Qualifiers
     source          1. .748
                     /organism="Aegilops speltoides"
                     /cultivar="F2 from 2-12-4-8-1-1-1-(1) x PI36909-12-811-(1
                     )"
                     /db_xref="taxon:4573"
                     /clone="WHE4211_F11_L20"
                     /clone_lib="Aegilops speltoides pre-meiotic anther cDNA
                     library"
                     /tissue_type="Anther"
                     /dev_stage="Pre-meiotic anthers"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth
                     chamber at the University of California, Davis (Akhunov in
                     Devorak's lab). Pre-meiotic anthers were harvested, total
                     RNA and poly(A) RNA were prepared, a cDNA library was made
                     , and the cDNA clones were in vivo excised to give
                     pBluescript phagemids at the University of California,
                     Davis (Akhunov in Devorak's lab). Plasmid DNA preparations
                     and DNA sequencing were performed in the OD Anderson lab
                     (all other authors)."
BASE COUNT      216 a    161 c    206 g    164 t      1 others
ORIGIN

  Query Match             2.5%;  Score 41.8;  DB 14;  Length 748;
  Best Local Similarity   52.6%;  Pred. No. 0.37;
  Matches   91;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy       1409 CTAACACTCAGGTTGAGAGTGGTAAAGGTGTTTTTGAGGGATGCTACTTCAACTACCCTG 1468
              || ||  | | | ||||  ||  ||    | |  |||||   || |  |  |  |  ||
Db        200 CTGACGATGAAGATGAGGATGAGAAGAAGGATACTGAGGAGGGCAAGGTTGAGGAAATTG 259
```

```
Qy       1469 ATGTTGACTTGAACAACTGGAAGAACGGTAAGTATGGTGCCTTGGAACTTTACTTTTTGG 1528
              |||  ||   | |  |   | ||||   ||| |      |  |||  |||    |  |
Db        260 ATGAAGAGAAGGAAGAAAAGGAGAAGAAAAAGAAGAAGATCAAGGAGGTTTCTCATGAGT 319

Qy       1529 GTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAGATC 1581
              | ||||||| ||    |   ||| |||  ||| || || | ||| | ||||||
Db        320 GGAACCTGATCAACAAGCAGAAGCCCATCTGGATGAGGAAGCCTGAGGAGATC 372
```

RESULT 11
BQ060219
LOCUS BQ060219 409 bp mRNA linear EST 01-APR-2002
DEFINITION Ts1168 Thellungiella salsuginea ZAP cDNA library Thellungiella salsuginea cDNA similar to reticuline oxidase - like protein, mRNA sequence.
ACCESSION BQ060219
VERSION BQ060219.1 GI:19855168
KEYWORDS EST.
SOURCE Thellungiella salsuginea.
ORGANISM Thellungiella salsuginea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Thellungiella.
REFERENCE 1 (bases 1 to 409)
AUTHORS Wang,Z.L., Li,P.H., Sun,Y.F., Zhang,Q., Zhao,Y.X. and Zhang,H.
TITLE Expressed sequence tags from a halophyte Thellungiella salsuginea cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Hui Zhang
Key Laboratory of Plant Stress Research
The Biology Department of Shandong Normal University
No.88, Wenhua East Road, Jinan, Shandong Province, 250014, PRC
Tel: (86)531-2960864
Fax: (86)531-2966954
Email: zhangh@sdnu.edu.cn.
FEATURES Location/Qualifiers
source 1. .409
/organism="Thellungiella salsuginea"
/db_xref="taxon:72664"
/clone_lib="Thellungiella salsuginea ZAP cDNA library"
/dev_stage="seedling"
/note="Organ: aerial part tissue; Vector: lambda zap; Site_1: EcoRI; Site_2: XhoI; total RNA extraction from NaCl(200mM) treated Thellungiella salsuginea by RNAgent kit(Promega); mRNA isolation by MESSAGEMAKER kit(GIBCO BRL); directional cDNA synthesis(EcoRI XhoI) by cDNA synthesis kit(STRATAGEN); the ZAP express library by GigapackIII Gold Cloning kit(STRATAGENE)"
BASE COUNT 92 a 110 c 108 g 99 t
ORIGIN

Query Match 2.5%; Score 41.4; DB 14; Length 409;
Best Local Similarity 53.4%; Pred. No. 0.3;
Matches 87; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```
Qy        357 GGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTA 416
              || || ||'|   | || |   | |  | |       | || || |  ||    |||
Db        205 GGTTTACTACCGAATCTGGGAGAAAAGCAAAATCCATGGATTCCCCGCCGGAGTTTGTCC 264

Qy        417 CTCCGTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCA 476
                | |||||| | || |||||  |   ||| || ||  |||||| | ||   ||   |
Db        265 GACGGTCGGTGTCGGAGGTCATTTAAGCGGCGGTGGGTACGGTAATATGCTGAGGAAGTT 324

Qy        477 CGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTT 519
              ||| ||| | ||||||   ||   || ||   | | ||||
Db        325 CGGATTGTCTGTCGATAACTTGATCGATGCGAAGATCGTCGAT 367
```

RESULT 12
CNS00LO0/c
LOCUS CNS00LO0 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR32D23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL068607
VERSION AL068607.1 GI:4958689
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR32D23"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT 155 a 166 c 7 g 284 t 489 others
ORIGIN

Query Match 2.5%; Score 41.4; DB 17; Length 1101;
Best Local Similarity 16.4%; Pred. No. 0.67;
Matches 68; Conservative 161; Mismatches 186; Indels 0; Gaps 0;

```
Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              :| :| :|:  ::|:    :|::|::||:::|:       | ::::   |:|:::   :|
Db       1014 DTTKGAKTKTTKKGKTNWTKTKKCKKTGKRKGDAKKAAAAADDKKTDKARAKKKGKKKGR 955

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              :::   :  : |::||: :   |: ::::| :: : :|:: :   |  : ::   :  :
Db        954 KWKKGGKGKKKTKDGTKKKKTTGKWDDDDAWWDWWADTDWKKAGGGGGKGKDAWTDTWDW 895

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              :   :: :: : :  :::    |:|:     :::| || |:, :  :|:: :|::|: :
Db        894 KDAWKKGDKKDTKKKKKKGGGKAKGRAAAAADRKGAAADGDAAAKTKKTKKAKTWKGKKK 835

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
                :  :|||: :: :|  ::|  |: :: ::: :       :|: :  |:   :  | ::
Db        834 GGGWGRKGTGRGKKGKGGTRKAGDTKKKKGRKKKKGTGDKKRTWWKTTTKTTGKTGTAWK 775

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
                ::: :    | :|:::   : :| |    :   :  : :|: :  ::|:: | :
Db        774 TKTGKKKKKGKKKGKKGKDWTKGTKWDTKTKTKGKKTGSGGGKKKTKYKGTKKTWWAAWK 715

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              |:  :  ::|  |:  :: ::: :: |:  :  :| :  ::|  :|: ::::
Db        714 AKKTKKTTKKKGTGAKTADKTKKKKKKDKGKTTGKWKKKTGKAAWWGTGDTRTKDKKKKA 655

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGT 835
              :| : : |:|:|  |::    |:| ::      |  |: ::|: :||:|::  |||
Db        654 WAAWNDTTDGKTKAKKADTKADTGKKGAKATKAAKAKGWDADADGTDTKKAKTGT 600
```

```
RESULT 13
AV805562/c
LOCUS       AV805562                406 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AV805562 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-43-G11 3',
            mRNA sequence.
ACCESSION   AV805562
VERSION     AV805562.1  GI:19839547
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES                Location/Qualifiers
     source             1. .406
                        /organism="Arabidopsis thaliana"
                        /db_xref="taxon:3702"
                        /clone="RAFL09-43-G11"
                        /clone_lib="RAFL9"
                        /dev_stage="plants at various developmental stages from
                        germination to mature seeds"
                        /lab_host="DH10B"
                        /note="Site_1: BamHI; Site_2: SalI; subjected to
                        dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                        hr) treatments"
BASE COUNT      122 a     99 c     64 g    121 t
ORIGIN

  Query Match             2.5%;  Score 41;  DB 10;  Length 406;
  Best Local Similarity   61.9%;  Pred. No. 0.4;
  Matches   65;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       1519 TACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAG 1578
              || || ||||| ||  |||| ||||| ||  | |  |||     ||  |||||| |  |
Db        204 TATTTCTTGGGGAATTTGAAGAGATTAATGGATGTTAAAGCTAAGTATGATCCTGATAAT 145

Qy       1579 ATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAGGAG 1623
              ||||||,||| | |||  ||| ||| |    | | | ||| ||
Db        144 TTCTTCAAGAACGAGCAGAGTATTCCTCCGGTTCGTGTAAAGTAG 100


RESULT 14
BF202701
LOCUS       BF202701                416 bp    mRNA    linear   EST 06-NOV-2000
DEFINITION  WHE1777_C01_F01ZS Wheat pre-anthesis spike cDNA library Triticum
            aestivum cDNA clone WHE1777_C01_F01, mRNA sequence.
ACCESSION   BF202701
VERSION     BF202701.1  GI:11117443
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Pre-anthesis spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES                Location/Qualifiers
     source             1. .416
                        /organism="Triticum aestivum"
                        /cultivar="Chinese Spring"
                        /db_xref="taxon:4565"
                        /clone="WHE1777_C01_F01"
                        /clone_lib="Wheat pre-anthesis spike cDNA library"
                        /tissue_type="Spike before anthesis"
                        /dev_stage="Adult plant"
                        /lab_host="E. coli SOLR"
                        /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                        Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                        greenhouse.  Whole spike with awns trimmed,  white, green
                        and yellow anther were collected and total RNA, and
                        poly(A) RNA were prepared, a cDNA library was made, and
                        the cDNA  clones were in vivo excised to give pBluescript
                        phagemids in the TJ  Close lab (Choi, Close, Fenton) at
                        the University of California,  Riverside.  Plasmid DNA
                        preparations and DNA sequencing were performed  in the OD
                        Anderson lab (all other authors)."
BASE COUNT      131 a     78 c    125 g     82 t
ORIGIN

  Query Match             2.5%;  Score 41;  DB 12;  Length 416;
  Best Local Similarity   52.0%;  Pred. No. 0.41;
  Matches   92;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy       1409 CTAACACTCAGGTTGAGAGTGGTAAAGGTGTTTTTGAGGGATGCTACTTCAACTACCCTG 1468
              || ||  | | | ||||  ||  ||    | |  |||||   || |  |  |  |   ||
Db        214 CTGACGATGAAGATGAGGATGAGAAGAAGGATACTGAGGAGGGCAAGGTTGAGGAAATTG 273

Qy       1469 ATGTTGACTTGAACAACTGGAAGAACGGTAAGTATGGTGCCTTGGAACTTTACTTTTTGG 1528
              |||  ||   | |  |   | ||||   ||| |      |  |||  |||       |
Db        274 ATGAAGAGAAGGAAGAAAAGGAGAAGAAAAAGAAGAAGATCAAGGAGGTTTCTCACGAGT 333

Qy       1529 GTAACCTGAACAGATTGATCAAGGCCAAATGGTTGTGGGATCCTAACGAGATCTTCA 1585
              | ||||||| ||    |  |||  |||  ||| || || | ||| | ||||||  ||
Db        334 GGAACCTGATCAACAAGCAGAAGCCCATCTGGATGAGGAAGCCTGAGGAGATCACCA 390


RESULT 15
BU013165
LOCUS       BU013165                569 bp    mRNA    linear   EST 22-AUG-2002
DEFINITION  QGJ3O15.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGJ3O15, mRNA sequence.
ACCESSION   BU013165
VERSION     BU013165.1  GI:22447560
KEYWORDS    EST.
SOURCE      Lactuca sativa.
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
            Lactuca.
```

```
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig4252, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGJ3  row: O  column: 15.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Lactuca sativa"
                     /cultivar="L.serriola"
                     /db_xref="taxon:4236"
                     /clone="QGJ3O15"
                     /clone_lib="QG_EFGHJ lettuce serriola"
                     /lab_host="E.coli"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_LIB=QG_EFGHJ lettuce serriola
                     TAG_TISSUE=flowers pre-fertilized
                     TAG_SEQ=GCTTGACGGG"
BASE COUNT      166 a     81 c    131 g    191 t
ORIGIN

  Query Match              2.5%;  Score 41;  DB 14;  Length 569;
  Best Local Similarity   60.2%;  Pred. No. 0.52;
  Matches   68;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1302 GCAGTACCAGACATACTGGCAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAG 1361
              | | || |||||| |||      ||| |    |||| || |||| |      |  | |||
Db         51 GAAATATCAGACACACTTTTCTGAATACATAAAGGCTGGAGTTGAAGCAGACAGCATTGA 110

Qy       1362 AGACTTTTACGAGGAGATGTATGAGCCTTATGGTGGTGTTCCAGACCCTAACA 1414
              |||  | ||| || || || |||   ||  | ||| || |||| ||||||
Db        111 AGAGGTCTACAAGAAGGTGCATGCAGCTATTCGTGCTGATCCAACCCCTAAGA 163


Search completed: June 15, 2003, 03:39:35
Job time : 2465 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 00:41:27 ; Search time 388 Seconds
(without alignments)
9541.972 Million cell updates/sec

Title: US-09-998-284-1
Perfect score: 1644
Sequence: 1 atggctactttgccacaaaa.....ctaagcagactaaatagtag 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1644	100.0	1644	22	AAS06173	Hexose oxidase (HO
	2	1644	100.0	1644	22	AAF23590	Chondrus cripus he
	3	1632.8	99.3	1644	21	AAZ94039	Synthetic hexose o
	4	1279.2	77.8	1801	18	AAT76552	DNA encoding hexos
c	5	126	7.7	126	21	AAZ94026	Primer hox16 used
c	6	126	7.7	126	22	AAS06168	PCR primer 6 used
	7	120	7.3	120	21	AAZ94015	Primer hox5 used i
c	8	120	7.3	120	21	AAZ94024	Primer hox14 used
	9	120	7.3	120	22	AAS06157	PCR primer hox5a u
c	10	120	7.3	120	22	AAS06166	PCR primer 4 used
	11	118	7.2	118	21	AAZ94019	Primer hox9 used i
	12	118	7.2	118	22	AAS06161	PCR primer hox9a u
	13	117	7.1	117	21	AAZ94029	Primer hox19 used
	14	117	7.1	117	22	AAS06171	PCR primer 9 used
c	15	116.4	7.1	118	22	AAS06162	PCR primer hox10b
c	16	116	7.1	116	21	AAZ94018	Primer hox8 used i
c	17	116	7.1	116	22	AAS06160	PCR primer hox8b u
c	18	113.2	6.9	118	21	AAZ94020	Primer hox10 used
c	19	111	6.8	111	21	AAZ94028	Primer hox18 used
c	20	111	6.8	111	22	AAS06170	PCR primer 8 used
c	21	109	6.6	109	21	AAZ94016	Primer hox6 used i
	22	109	6.6	109	21	AAZ94017	Primer hox7 used i
c	23	109	6.6	109	22	AAS06158	PCR primer hox6b u
	24	109	6.6	109	22	AAS06159	PCR primer hox7a u
	25	108	6.6	108	21	AAZ94025	Primer hox15 used
	26	108	6.6	108	21	AAZ94027	Primer hox17 used
	27	108	6.6	108	21	AAZ94031	Primer BglII-NsiI
	28	108	6.6	108	22	AAS06167	PCR primer 5 used
	29	108	6.6	108	22	AAS06169	PCR primer 7 used
c	30	107	6.5	107	21	AAZ94012	Primer hox2 used i
c	31	107	6.5	107	22	AAS06154	PCR primer hox2B u
	32	106	6.4	106	22	AAS06153	PCR primer hox1A u
	33	106	6.4	106	22	AAS06155	PCR primer hox3a u
c	34	106	6.4	106	22	AAS06156	PCR primer hox4b u
c	35	104.4	6.4	106	21	AAZ94014	Primer hox4 used i
c	36	102	6.2	102	22	AAS06164	PCR primer 2 used
c	37	100	6.1	100	21	AAZ94032	Primer BglII-NsiI
c	38	98.8	6.0	102	21	AAZ94022	Primer hox12 used
	39	94	5.7	105	21	AAZ94013	Primer hox3 used i
	40	92.8	5.6	96	21	AAZ94021	Primer hox11 used
	41	92.8	5.6	96	22	AAS06163	PCR primer 1 used
	42	85.2	5.2	90	21	AAZ94023	Primer hox13 used
	43	85.2	5.2	90	22	AAS06165	PCR primer 3 used
	44	72.2	4.4	83	21	AAZ94037	Primer -SKL-sense
c	45	68.2	4.1	83	21	AAZ94038	Primer -SKL-antise

ALIGNMENTS

RESULT 1
AAS06173
ID AAS06173 standard; DNA; 1644 BP.
XX
AC AAS06173;
XX
DT 12-SEP-2001 (first entry)
XX
DE Hexose oxidase (HOX) synthetic gene sequence.
XX
KW HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
KW food manufacturing; beverage; detergent; baking; dough improving agent;
KW D-hexose:O2-oxireductase; ds.
XX
OS Chondrus crispus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1641
FT /*tag= a
FT /product= "Hexose reductase (HOX)"
FT /EC_number= "EC 1.1.3.5"
XX
PN WO200138544-A1.
XX
PD 31-MAY-2001.
XX
PF 24-NOV-2000; 2000WO-IB01886.
XX
PR 24-NOV-1999; 99GB-0027801.
XX
PA (DANI-) DANISCO AS.

XX
PI Johansen CL, Kjaerulff S, Madrid SM, Pedersen H, Poulsen CH;
PI Zargahi MR;
XX
DR WPI; 2001-367695/38.
DR P-PSDB; AAU02192.
XX
PT Releasing soluble or membrane associated intracellular protein from a
PT cell for manufacturing food, comprises contacting the cell with a
PT membrane extracting composition and causing the protein to be released
PT in soluble form -
XX
PS Claim 17; Fig 6; 108pp; English.
XX
CC The sequence represents the coding sequence of synthetic hexose oxidase
CC (D-hexose:O2-oxireductase, EC 1.1.3.5), also referred to as HOX, gene.
CC The native HOX gene was altered using site-directed mutagenesis in order
CC to match the codon usage to known codon preferences of biotechnologically
CC relevant yeasts, such as Pichia sp., to facilitate high level production
CC in these organisms. The invention involves a method for releasing a
CC soluble or membrane associated intracellular protein of interest (POI)
CC from a cell involving contacting a cell comprising a soluble or membrane
CC associated intracellular POI with a membrane extracting composition (I)
CC and causing the POI to be released from the cell in a soluble form. The
CC method is useful for releasing POI, such as an interleukin I receptor
CC antagonist (IL-1ra) which involves contacting a transformed cell
CC comprising IL-1ra with (I) and causing IL-1ra to be released from the
CC transformed cell, in a soluble form. The method is also useful for
CC screening mutated cells or transformed cells producing elevated levels of
CC intracellular POI. The method is used to release a POI for manufacturing
CC food products, such as beverages, preparation of detergents, and in
CC baking as a dough improving agent. The method obtains a fast, specific
CC and economically efficient extraction of a soluble or membrane associated
CC intracellular POI without the use of conventional cell disruption
CC techniques. The resulting cell extract contains less contaminating
CC intracellular DNA and is relatively free of cell wall fragments. The
CC intracellular POI can be recovered from a eukaryotic host organism such
CC as yeast, before glycosylation takes place. The method can be used to
CC prevent contact of intracellular POI with the extracellular growth
CC medium.
XX
SQ Sequence 1644 BP; 430 A; 337 C; 404 G; 473 T; 0 other;

Query Match 100.0%; Score 1644; DB 22; Length 1644;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240

Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300

Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260

Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380

Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500
```

```
Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620

Qy       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
             |||||||||||||||||||||||||
Db       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
```

RESULT 2
AAF23590
ID AAF23590 standard; DNA; 1644 BP.
XX
AC AAF23590;
XX
DT 23-MAR-2001 (first entry)
XX
DE Chondrus cripus hexose oxidase enzyme DNA.
XX
KW Hexose oxidase; marine alga; anti-fouling; ds.
XX
OS Chondrus cripus.
XX
PN WO200075293-A2.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-IB00829.
XX
PR 04-JUN-1999; 99GB-0013050.
XX
PA (DANI-) DANISCO AS.
XX
PI Poulsen CH, Kragh KM;
XX
DR WPI; 2001-112148/12.
XX
PT New anti-fouling composition, useful as a coating for treating
PT different surfaces, e.g. outdoor woodwork, external surface of a
PT central heating system, or a hull of a marine vessel -
XX
PS Claim 5; Page 35-36; 36pp; English.
XX
CC The present invention relates to a new anti-fouling composition. The
CC composition involves a surface coating material, a hexose oxidase
CC enzyme obtained from a marine organism and a substrate for the
CC enzyme. The anti-fouling composition is useful as a coating formulated
CC for treating a surface, e.g. outdoor wood work, external surface of a
CC central heating system, or a hull of a marine vessel. It is
CC also useful as an anti-fouling agent for marine structures exposed to
CC seawater flora and fauna.
XX
SQ Sequence 1644 BP; 430 A; 337 C; 404 G; 473 T; 0 other;

Query Match 100.0%; Score 1644; DB 22; Length 1644;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240

Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300

Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
```

```
Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380

Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500

Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620

Qy       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
             ||||||||||||||||||||||||
Db       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
```

RESULT 3
AAZ94039
ID AAZ94039 standard; DNA; 1644 BP.
XX
AC AAZ94039;
XX
DT 29-AUG-2000 (first entry)
XX
DE Synthetic hexose oxidase gene.
XX
KW Hexose oxidase; production; fermentation; synthetic; modification;
KW prokaryote; eukaryote; ss.
XX
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT CDS   1..1641
FT   /*tag= a
FT   /product= Hexose oxidase
XX
PN EP1008651-A2.
XX
PD 14-JUN-2000.
XX
PF 01-DEC-1999; 99EP-0204068.
XX
PR 09-DEC-1998; 98DK-0001630.
XX
PA (BIOT-) BIOTEKNOLOGISK INST.
XX
PI Stougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;
XX
DR WPI; 2000-389309/34.
DR P-PSDB; AAY83619.
XX
PT Nucleic acid fragment useful for producing large amounts of hexose
PT oxidase comprises nucleotide sequence encoding hexose oxidase that is
PT modified by at least one codon
XX
PS Claim 1; Page 25-27; 42pp; English.
XX
CC Wild type hexose oxidase can be produced by isolating a nucleotide
CC sequence coding for hexose oxidase naturally produced by an organism
CC and modifying the sequence so that the modified sequence is
CC expressed, under identical conditions, at a level at least 10% higher
CC than the non-modified sequence. The method is useful for producing
CC hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
CC of producing hexose oxidase were to isolate the enzyme from a source
CC that naturally produces the enzyme, using host organisms which
CC produce a relatively low level of expression. Therefore, industrial
CC production of the enzyme is not feasible using this method. This is a
CC much improved method which produces at least 250 mg enzyme per liter
CC of fermentation medium.
XX
SQ Sequence 1644 BP; 430 A; 344 C; 401 G; 469 T; 0 other;

Query Match 99.3%; Score 1632.8; DB 21; Length 1644;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1637; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240

Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATCATCAACGTTACTGGTTTGGTT 300

Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              |||||||||||||||||||||||||| ||||| ||||||||||||||||||||||||||||
Db        481 TTGCCAGTCGATTGGTTATCCGGTGTCGAAGTCGTCGTTAAGCCAGTCTTGACCGAAGAC 540

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
```

```
Qy       841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900
Db       841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900

Qy       901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960
Db       901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960

Qy       961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020
Db       961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020

Qy      1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080
Db      1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080

Qy      1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140
Db      1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140

Qy      1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200
Db      1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTACTGCCGAA 1200

Qy      1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
Db      1201 ATGAAGGATGCTCTTCTTCAGCTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260

Qy      1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320
Db      1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320

Qy      1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380
Db      1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380

Qy      1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440
Db      1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440

Qy      1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500
Db      1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGCTAAG 1500

Qy      1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560
Db      1501 TATGGTGCCTTGGAACTCTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560

Qy      1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620
Db      1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620

Qy      1621 GAGCCTAAGCAGACTAAATAGTAG 1644
Db      1621 GAGCCTAAGCAGACTAAATAGTAG 1644
```

RESULT 4
AAT76552
ID AAT76552 standard; DNA; 1801 BP.
XX
AC AAT76552;
XX
DT 12-SEP-1997 (first entry)
XX
DE DNA encoding hexose oxidase, an antimicrobial/antioxidant agent.
XX
KW Hexose oxidase; Chondrus crispus; marine algae; recombinant;
KW antimicrobial; antioxidant; food preparation; dairy product; starch;
KW beverage; animal feed; silage; sugar reduction; cosmetics; dental;
KW toothpaste; dough; lactone production; ss.
XX
OS Chondrus crispus.
XX
FH Key Location/Qualifiers
FT CDS 84..1724
FT /*tag= a
XX
PN WO9640935-A1.
XX
PD 19-DEC-1996.
XX
PF 04-JUN-1996; 96WO-DK00238.
XX
PR 07-JUN-1995; 95US-0476910.
XX
PA (BIOT-) BIOTEKNOLOGISK INST.
XX
PI Hansen OC, Stougaard P;
XX
DR WPI; 1997-052332/05.
DR P-PSDB; AAW20076.
XX
PT Recombinant production of polypeptide having hexose oxidase activity
PT - used in food preparations as antibacterial and antioxidant agent
XX
PS Claim 28; Page 103-104; 124pp; English.
XX
CC AAT76552 encodes hexose oxidase (HO) of the marine algae species
CC Chondrus crispus. HO is useful in the production of food products, e.g.
CC dairy products, starch-containing food products (dough) and non-dairy
CC beverages. HO and active peptide fragments have antimicrobial and
CC antioxidant properties and act by removing all the oxygen in a food
CC packaging. HO and peptides of HO can also be used in an animal feed,
CC especially silage. Further uses are to reduce, or analyse, the sugar
CC content in a food, in the production of cosmetics, tooth care products
CC or a pharmaceutical product and in lactone production. HO can be
CC recombinantly produced in industrially appropriate quantities, and at
CC a quality and purity level which renders the polypeptide suitable for
CC industrial purposes.
XX
SQ Sequence 1801 BP; 441 A; 448 C; 492 G; 420 T; 0 other;

Query Match 77.8%; Score 1279.2; DB 18; Length 1801;
Best Local Similarity 86.1%; Pred. No. 0;
Matches 1416; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

```
Qy         1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
Db        84 ATGGCTACTCTTCCTCAGAAAGACCCCGGTTATATTGTAATTGATGTCAACGCGGGCACC 143

Qy        61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
Db       144 GCGGACAAGCCGGACCCACGTCTCCCCTCCATGAAGCAGGGCTTCAACCGCCGCTGGATT 203

Qy       121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
Db       204 GGAACTAATATCGATTTCGTTTATGTCGTGTACACTCCTCAAGGTGCTTGTACTGCACTT 263

Qy       181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
Db       264 GACCGTGCTATGGAAAAGTGTTCTCCCGGTACAGTCAGGATCGTCTCTGGCGGCCATTGC 323

Qy       241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
Db       324 TACGAGGACTTCGTATTTGACGAATGCGTCAAGGCCATCATCAACGTCACTGGTCTCGTT 383

Qy       301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
Db       384 GAGAGTGGTTATGACGACGATAGGGGTTACTTCGTCAGCAGTGGAGATACAAATTGGGGC 443

Qy       361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
Db       444 TCCTTCAAGACCTTGTTCAGAGACCACGGAAGAGTTCTTCCCGGGGGTTCCTGCTACTCC 503

Qy       421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
```

```
Db        504 GTCGGCCTCGGTGGCCACATTGTCGGCGGAGGTGACGGCATTTTGGCCCGCTTGCATGGC 563

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540

Db        564 CTCCCCGTCGATTGGCTCAGCGGCGTGGAGGTCGTCGTTAAGCCAGTCCTCACCGAAGAC 623

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600

Db        624 TCGGTACTCAAGTATGTGCACAAAGATTCCGAAGGCAACGACGGGGAGCTCTTTTGGGCA 683

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660

Db        684 CACACAGGTGGCGGTGGCGGAAACTTTGGAATCATCACCAAATACTACTTCAAGGATTTG 743

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720

Db        744 CCCATGTCTCCACGGGGCGTCATCGCATCAAATTTACACTTCAGCTGGGACGGTTTCACG 803

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780

Db        804 AGAGATGCCTTGCAGGATTTGTTGACAAAGTACTTCAAACTTGCCAGATGTGATTGGAAG 863

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840

Db        864 AATACGGTTGGCAAGTTTCAAATCTTCCATCAGGCAGCGGAAGAGTTTGTCATGTACTTG 923

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900

Db        924 TATACATCCTACTCGAACGACGCCGAGCGCGAAGTTGCCCAAGACCGTCACTATCATTTG 983

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960

Db        984 GAGGCTGACATAGAACAGATCTACAAAACATGCGAGCCCACCAAAGCGCTTGGCGGGCAT 1043

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020

Db       1044 GCTGGGTGGGCGCCGTTCCCCGTGCGGCCGCGCAAGAGGCACACATCCAAGACGTCGTAT 1103

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080

Db       1104 ATGCATGACGAGACGATGGACTACCCCTTCTACGCGCTCACTGAGACGATCAACGGCTCC 1163

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140

Db       1164 GGGCCGAATCAGCGCGGCAAGTACAAGTCTGCGTACATGATCAAGGATTTCCCGGATTTC 1223

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200

Db       1224 CAGATCGACGTGATCTGGAAATACCTTACGGAGGTCCCGGACGGCTTGACTAGTGCCGAA 1283

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260

Db       1284 ATGAAGGATGCCTTACTCCAGGTGGACATGTTTGGTGGTGAGATTCACAAGGTGGTCTGG 1343

Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320

Db       1344 GATGCGACGGCAGTCGCGCAGCGCGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1403

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380

Db       1404 CAGGAAGAAGACAAGGATGCAGTGAACCTCAAGTGGATTAGAGACTTTTACGAGGAGATG 1463

Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440

Db       1464 TATGAGCCGTATGGCGGGGTTCCAGACCCCAACACGCAGGTGGAGAGTGGTAAAGGTGTG 1523

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500

Db       1524 TTTGAGGGATGCTACTTCAACTACCCGGATGTGGACTTGAACAACTGGAAGAACGGCAAG 1583

Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560

Db       1584 TATGGTGCCCTCGAACTTTACTTTTTGGGTAACCTGAACCGCCTCATCAAGGCCAAATGG 1643

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620

Db       1644 TTGTGGGATCCCAACGAGATCTTCACAAACAAACAGAGCATCCCTACTAAACCTCTTAAG 1703

Qy       1621 GAGCCTAAGCAGACTAAATAGTAG 1644

Db       1704 GAGCCCAAGCAGACGAAATAGTAG 1727
```

RESULT 5
AAZ94026/c
ID AAZ94026 standard; DNA; 126 BP.
XX
AC AAZ94026;
XX
DT 29-AUG-2000 (first entry)
XX
DE Primer hox16 used in production of synthetic hexose oxidase gene.
XX
KW Hexose oxidase; production; fermentation; synthetic; modification;
KW prokaryote; eukaryote; ss.
XX
OS Synthetic.
XX
PN EP1008651-A2.
XX
PD 14-JUN-2000.
XX
PF 01-DEC-1999; 99EP-0204068.
XX
PR 09-DEC-1998; 98DK-0001630.
XX
PA (BIOT-) BIOTEKNOLOGISK INST.
XX
PI Stougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;
XX
DR WPI; 2000-389309/34.
XX
PT Nucleic acid fragment useful for producing large amounts of hexose
PT oxidase comprises nucleotide sequence encoding hexose oxidase that is
PT modified by at least one codon
XX
PS Example 2; Page 22; 42pp; English.
XX
CC Wild type hexose oxidase can be produced by isolating a nucleotide
CC sequence coding for hexose oxidase naturally produced by an organism
CC and modifying the sequence so that the modified sequence is
CC expressed, under identical conditions, at a level at least 10% higher
CC than the non-modified sequence. The method is useful for producing
CC hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
CC of producing hexose oxidase were to isolate the enzyme from a source
CC that naturally produces the enzyme, using host organisms which
CC produce a relatively low level of expression. Therefore, industrial
CC production of the enzyme is not feasible using this method. This is a
CC much improved method which produces at least 250 mg enzyme per liter
CC of fermentation medium. Twenty-eight primers (See AAZ94011-Z94038)
CC with codon usage possibly optimised for expression in P. pastoris
CC were used to produce the synthetic hexose oxidase gene (See AAZ94039).
XX
SQ Sequence 126 BP; 33 A; 35 C; 23 G; 35 T; 0 other;

Query Match 7.7%; Score 126; DB 21; Length 126;
Best Local Similarity 100.0%; Pred. No. 8.5e-28;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1210 GCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGGGATGCTACT 1269

Db        126 GCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGGGATGCTACT 67

Qy       1270 GCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGGCAGGAAGAA 1329
```

```
Db         66 GCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGGCAGGAAGAA 7

Qy       1330 GACAAG 1335
              ||||||
Db          6 GACAAG 1

RESULT 6
AAS06168/c
ID   AAS06168 standard; DNA; 126 BP.
XX
AC   AAS06168;
XX
DT   12-SEP-2001  (first entry)
XX
DE   PCR primer 6 used to clone synthetic HOX gene.
XX
KW   HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
KW   food manufacturing; beverage; detergent; baking; dough improving agent;
KW   D-hexose:O2-oxireductase; EC 1.1.3.5; PCR primer; ss.
XX
OS   Chondrus crispus.
OS   Synthetic.
XX
PN   WO200138544-A1.
XX
PD   31-MAY-2001.
XX
PF   24-NOV-2000; 2000WO-IB01886.
XX
PR   24-NOV-1999;   99GB-0027801.
XX
PA   (DANI-) DANISCO AS.
XX
PI   Johansen CL,  Kjaerulff S,  Madrid SM,  Pedersen H,  Poulsen CH;
PI   Zargahi MR;
XX
DR   WPI; 2001-367695/38.
XX
PT   Releasing soluble or membrane associated intracellular protein from a
PT   cell for manufacturing food, comprises contacting the cell with a
PT   membrane extracting composition and causing the protein to be released
PT   in soluble form  -
XX
PS   Claim 19; Fig 5; 108pp; English.
XX
CC   The sequence represents the PCR primer 6 used to clone synthetic
CC   hexose oxidase (D-hexose:O2-oxireductase, EC 1.1.3.5), (also referred to
CC   as HOX), gene. The native HOX gene was altered using site-directed
CC   mutagenesis in order to match the codon usage to known codon preferences
CC   of biotechnologically relevant yeasts, such as Pichia sp., to facilitate
CC   high level production in these organisms. The invention involves a method
CC   for releasing a soluble or membrane associated intracellular protein of
CC   interest (POI) from a cell involving contacting a cell comprising a
CC   soluble or membrane associated intracellular POI with a membrane
CC   extracting composition (I) and causing the POI to be released from the
CC   cell in a soluble form. The method is useful for releasing POI, such as
CC   an interleukin I receptor antagonist (IL-1ra) which involves contacting a
CC   transformed cell comprising IL-1ra with (I) and causing IL-1ra to be
CC   released from the transformed cell, in a soluble form. The method is also
CC   useful for screening mutated cells or transformed cells producing
CC   elevated levels of intracellular POI. The method is used to release a POI
CC   for manufacturing food products, such as beverages, preparation of
CC   detergents, and in baking as a dough improving agent. The method obtains
CC   a fast, specific and economically efficient extraction of a soluble or
CC   membrane associated intracellular POI without the use of conventional
CC   cell disruption techniques. The resulting cell extract contains less
CC   contaminating intracellular DNA and is relatively free of cell wall
CC   fragments. The intracellular POI can be recovered from a eukaryotic host
CC   organism such as yeast, before glycosylation takes place. The method can
CC   be used to prevent contact of intracellular POI with the extracellular
CC   growth medium.
XX

SQ   Sequence 126 BP; 33 A; 35 C; 23 G; 35 T; 0 other;

  Query Match             7.7%;  Score 126;  DB 22;  Length 126;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-28;
  Matches  126;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1210 GCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGGGATGCTACT 1269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          126 GCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGGGATGCTACT 67

Qy       1270 GCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGGCAGGAAGAA 1329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           66 GCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGGCAGGAAGAA 7

Qy       1330 GACAAG 1335
              ||||||
Db           6 GACAAG 1


RESULT 7
AAZ94015
ID   AAZ94015 standard; DNA; 120 BP.
XX
AC   AAZ94015;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Primer hox5 used in production of synthetic hexose oxidase gene.
XX
KW   Hexose oxidase; production; fermentation; synthetic; modification;
KW   prokaryote; eukaryote; ss.
XX
OS   Synthetic.
XX
PN   EP1008651-A2.
XX
PD   14-JUN-2000.
XX
PF   01-DEC-1999;   99EP-0204068.
XX
PR   09-DEC-1998;   98DK-0001630.
XX
PA   (BIOT-) BIOTEKNOLOGISK INST.
XX
PI   Stougaard P,  Pedersen LH,  Wolff AM,  Poulsen U,  Hansen OC;
XX
DR   WPI; 2000-389309/34.
XX
PT   Nucleic acid fragment useful for producing large amounts of hexose
PT   oxidase comprises nucleotide sequence encoding hexose oxidase that is
PT   modified by at least one codon
XX
PS   Example 2; Page 20; 42pp; English.
XX
CC   Wild type hexose oxidase can be produced by isolating a nucleotide
CC   sequence coding for hexose oxidase naturally produced by an organism
CC   and modifying the sequence so that the modified sequence is
CC   expressed, under identical conditions, at a level at least 10% higher
CC   than the non-modified sequence. The method is useful for producing
CC   hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
CC   of producing hexose oxidase were to isolate the enzyme from a source
CC   that naturally produces the enzyme, using host organisms which
CC   produce a relatively low level of expression. Therefore, industrial
CC   production of the enzyme is not feasible using this method. This is a
CC   much improved method which produces at least 250 mg enzyme per liter
CC   of fermentation medium. Twenty-eight primers (See AAZ94011-Z94038)
CC   with codon usage possibly optimised for expression in P. pastoris
CC   were used to produce the synthetic hexose oxidase gene (See AAZ94039).
XX
SQ   Sequence 120 BP; 26 A; 26 C; 31 G; 37 T; 0 other;

  Query Match             7.3%;  Score 120;  DB 21;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-26;
```

```
  Matches  120;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        280 ATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCT 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCT 60

Qy        340 TCCGGTGACACCAACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTG 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCGGTGACACCAACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTG 120


RESULT 8
AAZ94024/c
ID   AAZ94024 standard; DNA; 120 BP.
XX
AC   AAZ94024;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Primer hox14 used in production of synthetic hexose oxidase gene.
XX
KW   Hexose oxidase; production; fermentation; synthetic; modification;
KW   prokaryote; eukaryote; ss.
XX
OS   Synthetic.
XX
PN   EP1008651-A2.
XX
PD   14-JUN-2000.
XX
PF   01-DEC-1999;   99EP-0204068.
XX
PR   09-DEC-1998;   98DK-0001630.
XX
PA   (BIOT-) BIOTEKNOLOGISK INST.
XX
PI   Stougaard P,  Pedersen LH,  Wolff AM,  Poulsen U,  Hansen OC;
XX
DR   WPI; 2000-389309/34.
XX
PT   Nucleic acid fragment useful for producing large amounts of hexose
PT   oxidase comprises nucleotide sequence encoding hexose oxidase that is
PT   modified by at least one codon
XX
PS   Example 2; Page 22; 42pp; English.
XX
CC   Wild type hexose oxidase can be produced by isolating a nucleotide
CC   sequence coding for hexose oxidase naturally produced by an organism
CC   and modifying the sequence so that the modified sequence is
CC   expressed, under identical conditions, at a level at least 10% higher
CC   than the non-modified sequence. The method is useful for producing
CC   hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
CC   of producing hexose oxidase were to isolate the enzyme from a source
CC   that naturally produces the enzyme, using host organisms which
CC   produce a relatively low level of expression. Therefore, industrial
CC   production of the enzyme is not feasible using this method. This is a
CC   much improved method which produces at least 250 mg enzyme per liter
CC   of fermentation medium. Twenty-eight primers (See AAZ94011-Z94038)
CC   with codon usage possibly optimised for expression in P. pastoris
CC   were used to produce the synthetic hexose oxidase gene (See AAZ94039).
XX
SQ   Sequence 120 BP; 33 A; 25 C; 29 G; 33 T; 0 other;

  Query Match             7.3%;  Score 120;  DB 21;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-26;
  Matches  120;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1027 GACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCCGGTCCT 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 GACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCCGGTCCT 61

Qy       1087 AATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTCCAGATT 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          60 AATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTCCAGATT 1


RESULT 9
AAS06157
ID   AAS06157 standard; DNA; 120 BP.
XX
AC   AAS06157;
XX
DT   12-SEP-2001  (first entry)
XX
DE   PCR primer hox5a used to clone synthetic HOX gene.
XX
KW   HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
KW   food manufacturing; beverage; detergent; baking; dough improving agent;
KW   D-hexose:O2-oxireductase; EC 1.1.3.5; PCR primer; hox5a; ss.
XX
OS   Chondrus crispus.
OS   Synthetic.
XX
PN   WO200138544-A1.
XX
PD   31-MAY-2001.
XX
PF   24-NOV-2000; 2000WO-IB01886.
XX
PR   24-NOV-1999;   99GB-0027801.
XX
PA   (DANI-) DANISCO AS.
XX
PI   Johansen CL,  Kjaerulff S,  Madrid SM,  Pedersen H,  Poulsen CH;
PI   Zargahi MR;
XX
DR   WPI; 2001-367695/38.
XX
PT   Releasing soluble or membrane associated intracellular protein from a
PT   cell for manufacturing food, comprises contacting the cell with a
PT   membrane extracting composition and causing the protein to be released
PT   in soluble form  -
XX
PS   Claim 19; Fig 5; 108pp; English.
XX
CC   The sequence represents the PCR primer hox5a used to clone synthetic
CC   hexose oxidase (D-hexose:O2-oxireductase, EC 1.1.3.5), (also referred to
CC   as HOX), gene. The native HOX gene was altered using site-directed
CC   mutagenesis in order to match the codon usage to known codon preferences
CC   of biotechnologically relevant yeasts, such as Pichia sp., to facilitate
CC   high level production in these organisms. The invention involves a method
CC   for releasing a soluble or membrane associated intracellular protein of
CC   interest (POI) from a cell involving contacting a cell comprising a
CC   soluble or membrane associated intracellular POI with a membrane
CC   extracting composition (I) and causing the POI to be released from the
CC   cell in a soluble form. The method is useful for releasing POI, such as
CC   an interleukin I receptor antagonist (IL-1ra) which involves contacting a
CC   transformed cell comprising IL-1ra with (I) and causing IL-1ra to be
CC   released from the transformed cell, in a soluble form. The method is also
CC   useful for screening mutated cells or transformed cells producing
CC   elevated levels of intracellular POI. The method is used to release a POI
CC   for manufacturing food products, such as beverages, preparation of
CC   detergents, and in baking as a dough improving agent. The method obtains
CC   a fast, specific and economically efficient extraction of a soluble or
CC   membrane associated intracellular POI without the use of conventional
CC   cell disruption techniques. The resulting cell extract contains less
CC   contaminating intracellular DNA and is relatively free of cell wall
CC   fragments. The intracellular POI can be recovered from a eukaryotic host
CC   organism such as yeast, before glycosylation takes place. The method can
CC   be used to prevent contact of intracellular POI with the extracellular
CC   growth medium.
XX
SQ   Sequence 120 BP; 26 A; 26 C; 31 G; 37 T; 0 other;

  Query Match             7.3%;  Score 120;  DB 22;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-26;
```

```
  Matches  120;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        280 ATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCT 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCT 60

Qy        340 TCCGGTGACACCAACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTG 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCGGTGACACCAACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTG 120
```

RESULT 10
AAS06166/c
ID AAS06166 standard; DNA; 120 BP.
XX
AC AAS06166;
XX
DT 12-SEP-2001 (first entry)
XX
DE PCR primer 4 used to clone synthetic HOX gene.
XX
KW HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
KW food manufacturing; beverage; detergent; baking; dough improving agent;
KW D-hexose:O2-oxireductase; EC 1.1.3.5; PCR primer; ss.
XX
OS Chondrus crispus.
OS Synthetic.
XX
PN WO200138544-A1.
XX
PD 31-MAY-2001.
XX
PF 24-NOV-2000; 2000WO-IB01886.
XX
PR 24-NOV-1999; 99GB-0027801.
XX
PA (DANI-) DANISCO AS.
XX
PI Johansen CL, Kjaerulff S, Madrid SM, Pedersen H, Poulsen CH;
PI Zargahi MR;
XX
DR WPI; 2001-367695/38.
XX
PT Releasing soluble or membrane associated intracellular protein from a
PT cell for manufacturing food, comprises contacting the cell with a
PT membrane extracting composition and causing the protein to be released
PT in soluble form -
XX
PS Claim 19; Fig 5; 108pp; English.
XX
CC The sequence represents the PCR primer 4 used to clone synthetic
CC hexose oxidase (D-hexose:O2-oxireductase, EC 1.1.3.5), (also referred to
CC as HOX), gene. The native HOX gene was altered using site-directed
CC mutagenesis in order to match the codon usage to known codon preferences
CC of biotechnologically relevant yeasts, such as Pichia sp., to facilitate
CC high level production in these organisms. The invention involves a method
CC for releasing a soluble or membrane associated intracellular protein of
CC interest (POI) from a cell involving contacting a cell comprising a
CC soluble or membrane associated intracellular POI with a membrane
CC extracting composition (I) and causing the POI to be released from the
CC cell in a soluble form. The method is useful for releasing POI, such as
CC an interleukin I receptor antagonist (IL-1ra) which involves contacting a
CC transformed cell comprising IL-1ra with (I) and causing IL-1ra to be
CC released from the transformed cell, in a soluble form. The method is also
CC useful for screening mutated cells or transformed cells producing
CC elevated levels of intracellular POI. The method is used to release a POI
CC for manufacturing food products, such as beverages, preparation of
CC detergents, and in baking as a dough improving agent. The method obtains
CC a fast, specific and economically efficient extraction of a soluble or
CC membrane associated intracellular POI without the use of conventional
CC cell disruption techniques. The resulting cell extract contains less
CC contaminating intracellular DNA and is relatively free of cell wall
CC fragments. The intracellular POI can be recovered from a eukaryotic host
CC organism such as yeast, before glycosylation takes place. The method can
CC be used to prevent contact of intracellular POI with the extracellular
CC growth medium.
XX
SQ Sequence 120 BP; 33 A; 25 C; 29 G; 33 T; 0 other;

```
  Query Match             7.3%;  Score 120;  DB 22;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-26;
  Matches  120;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1027 GACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCCGGTCCT 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 GACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCCGGTCCT 61

Qy       1087 AATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTCCAGATT 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 AATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTCCAGATT 1
```

RESULT 11
AAZ94019
ID AAZ94019 standard; DNA; 118 BP.
XX
AC AAZ94019;
XX
DT 29-AUG-2000 (first entry)
XX
DE Primer hox9 used in production of synthetic hexose oxidase gene.
XX
KW Hexose oxidase; production; fermentation; synthetic; modification;
KW prokaryote; eukaryote; ss.
XX
OS Synthetic.
XX
PN EP1008651-A2.
XX
PD 14-JUN-2000.
XX
PF 01-DEC-1999; 99EP-0204068.
XX
PR 09-DEC-1998; 98DK-0001630.
XX
PA (BIOT-) BIOTEKNOLOGISK INST.
XX
PI Stougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;
XX
DR WPI; 2000-389309/34.
XX
PT Nucleic acid fragment useful for producing large amounts of hexose
PT oxidase comprises nucleotide sequence encoding hexose oxidase that is
PT modified by at least one codon
XX
PS Example 2; Page 21; 42pp; English.
XX
CC Wild type hexose oxidase can be produced by isolating a nucleotide
CC sequence coding for hexose oxidase naturally produced by an organism
CC and modifying the sequence so that the modified sequence is
CC expressed, under identical conditions, at a level at least 10% higher
CC than the non-modified sequence. The method is useful for producing
CC hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
CC of producing hexose oxidase were to isolate the enzyme from a source
CC that naturally produces the enzyme, using host organisms which
CC produce a relatively low level of expression. Therefore, industrial
CC production of the enzyme is not feasible using this method. This is a
CC much improved method which produces at least 250 mg enzyme per liter
CC of fermentation medium. Twenty-eight primers (See AAZ94011-Z94038)
CC with codon usage possibly optimised for expression in P. pastoris
CC were used to produce the synthetic hexose oxidase gene (See AAZ94039).
XX
SQ Sequence 118 BP; 30 A; 27 C; 22 G; 39 T; 0 other;

```
  Query Match             7.2%;  Score 118;  DB 21;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-25;
```

```
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        639 CAAATACTACTTCAAGGATTTGCCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACA 698
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAATACTACTTCAAGGATTTGCCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACA 60

Qy        699 CTTCTCTTGGGACGGTTTCACTAGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTC 756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTTCTCTTGGGACGGTTTCACTAGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTC 118


RESULT 12
AAS06161
ID   AAS06161 standard; DNA; 118 BP.
XX
AC   AAS06161;
XX
DT   12-SEP-2001  (first entry)
XX
DE   PCR primer hox9a used to clone synthetic HOX gene.
XX
KW   HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
KW   food manufacturing; beverage; detergent; baking; dough improving agent;
KW   D-hexose:O2-oxireductase; EC 1.1.3.5; PCR primer; hox9a; ss.
XX
OS   Chondrus crispus.
OS   Synthetic.
XX
PN   WO200138544-A1.
XX
PD   31-MAY-2001.
XX
PF   24-NOV-2000; 2000WO-IB01886.
XX
PR   24-NOV-1999;   99GB-0027801.
XX
PA   (DANI-) DANISCO AS.
XX
PI   Johansen CL,  Kjaerulff S,  Madrid SM,  Pedersen H,  Poulsen CH;
PI   Zargahi MR;
XX
DR   WPI; 2001-367695/38.
XX
PT   Releasing soluble or membrane associated intracellular protein from a
PT   cell for manufacturing food, comprises contacting the cell with a
PT   membrane extracting composition and causing the protein to be released
PT   in soluble form  -
XX
PS   Claim 19; Fig 5; 108pp; English.
XX
CC   The sequence represents the PCR primer hox9a used to clone synthetic
CC   hexose oxidase (D-hexose:O2-oxireductase, EC 1.1.3.5), (also referred to
CC   as HOX), gene. The native HOX gene was altered using site-directed
CC   mutagenesis in order to match the codon usage to known codon preferences
CC   of biotechnologically relevant yeasts, such as Pichia sp., to facilitate
CC   high level production in these organisms. The invention involves a method
CC   for releasing a soluble or membrane associated intracellular protein of
CC   interest (POI) from a cell involving contacting a cell comprising a
CC   soluble or membrane associated intracellular POI with a membrane
CC   extracting composition (I) and causing the POI to be released from the
CC   cell in a soluble form. The method is useful for releasing POI, such as
CC   an interleukin I receptor antagonist (IL-1ra) which involves contacting a
CC   transformed cell comprising IL-1ra with (I) and causing IL-1ra to be
CC   released from the transformed cell, in a soluble form. The method is also
CC   useful for screening mutated cells or transformed cells producing
CC   elevated levels of intracellular POI. The method is used to release a POI
CC   for manufacturing food products, such as beverages, preparation of
CC   detergents, and in baking as a dough improving agent. The method obtains
CC   a fast, specific and economically efficient extraction of a soluble or
CC   membrane associated intracellular POI without the use of conventional
CC   cell disruption techniques. The resulting cell extract contains less
CC   contaminating intracellular DNA and is relatively free of cell wall
CC   fragments. The intracellular POI can be recovered from a eukaryotic host
CC   organism such as yeast, before glycosylation takes place. The method can
CC   be used to prevent contact of intracellular POI with the extracellular
CC   growth medium.
XX
SQ   Sequence 118 BP; 30 A; 27 C; 22 G; 39 T; 0 other;

  Query Match             7.2%;  Score 118;  DB 22;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-25;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        639 CAAATACTACTTCAAGGATTTGCCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACA 698
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAATACTACTTCAAGGATTTGCCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACA 60

Qy        699 CTTCTCTTGGGACGGTTTCACTAGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTC 756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTTCTCTTGGGACGGTTTCACTAGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTC 118


RESULT 13
AAZ94029
ID   AAZ94029 standard; DNA; 117 BP.
XX
AC   AAZ94029;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Primer hox19 used in production of synthetic hexose oxidase gene.
XX
KW   Hexose oxidase; production; fermentation; synthetic; modification;
KW   prokaryote; eukaryote; ss.
XX
OS   Synthetic.
XX
PN   EP1008651-A2.
XX
PD   14-JUN-2000.
XX
PF   01-DEC-1999;   99EP-0204068.
XX
PR   09-DEC-1998;   98DK-0001630.
XX
PA   (BIOT-) BIOTEKNOLOGISK INST.
XX
PI   Stougaard P,  Pedersen LH,  Wolff AM,  Poulsen U,  Hansen OC;
XX
DR   WPI; 2000-389309/34.
XX
PT   Nucleic acid fragment useful for producing large amounts of hexose
PT   oxidase comprises nucleotide sequence encoding hexose oxidase that is
PT   modified by at least one codon
XX
PS   Example 2; Page 23; 42pp; English.
XX
CC   Wild type hexose oxidase can be produced by isolating a nucleotide
CC   sequence coding for hexose oxidase naturally produced by an organism
CC   and modifying the sequence so that the modified sequence is
CC   expressed, under identical conditions, at a level at least 10% higher
CC   than the non-modified sequence. The method is useful for producing
CC   hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
CC   of producing hexose oxidase were to isolate the enzyme from a source
CC   that naturally produces the enzyme, using host organisms which
CC   produce a relatively low level of expression. Therefore, industrial
CC   production of the enzyme is not feasible using this method. This is a
CC   much improved method which produces at least 250 mg enzyme per liter
CC   of fermentation medium. Twenty-eight primers (See AAZ94011-Z94038)
CC   with codon usage possibly optimised for expression in P. pastoris
CC   were used to produce the synthetic hexose oxidase gene (See AAZ94039).
XX
SQ   Sequence 117 BP; 36 A; 23 C; 26 G; 32 T; 0 other;

  Query Match             7.1%;  Score 117;  DB 21;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-25;
```

```
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1489 AAGAACGGTAAGTATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATC 1548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGAACGGTAAGTATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATC 60

Qy       1549 AAGGCCAAATGGTTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCT 1605
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGGCCAAATGGTTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCT 117


RESULT 14
AAS06171
ID   AAS06171 standard; DNA; 117 BP.
XX
AC   AAS06171;
XX
DT   12-SEP-2001  (first entry)
XX
DE   PCR primer 9 used to clone synthetic HOX gene.
XX
KW   HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
KW   food manufacturing; beverage; detergent; baking; dough improving agent;
KW   D-hexose:O2-oxireductase; EC 1.1.3.5; PCR primer; ss.
XX
OS   Chondrus crispus.
OS   Synthetic.
XX
PN   WO200138544-A1.
XX
PD   31-MAY-2001.
XX
PF   24-NOV-2000; 2000WO-IB01886.
XX
PR   24-NOV-1999;   99GB-0027801.
XX
PA   (DANI-) DANISCO AS.
XX
PI   Johansen CL,  Kjaerulff S,  Madrid SM,  Pedersen H,  Poulsen CH;
PI   Zargahi MR;
XX
DR   WPI; 2001-367695/38.
XX
PT   Releasing soluble or membrane associated intracellular protein from a
PT   cell for manufacturing food, comprises contacting the cell with a
PT   membrane extracting composition and causing the protein to be released
PT   in soluble form  -
XX
PS   Claim 19; Fig 5; 108pp; English.
XX
CC   The sequence represents the PCR primer 9 used to clone synthetic
CC   hexose oxidase (D-hexose:O2-oxireductase, EC 1.1.3.5), (also referred to
CC   as HOX), gene. The native HOX gene was altered using site-directed
CC   mutagenesis in order to match the codon usage to known codon preferences
CC   of biotechnologically relevant yeasts, such as Pichia sp., to facilitate
CC   high level production in these organisms. The invention involves a method
CC   for releasing a soluble or membrane associated intracellular protein of
CC   interest (POI) from a cell involving contacting a cell comprising a
CC   soluble or membrane associated intracellular POI with a membrane
CC   extracting composition (I) and causing the POI to be released from the
CC   cell in a soluble form. The method is useful for releasing POI, such as
CC   an interleukin I receptor antagonist (IL-1ra) which involves contacting a
CC   transformed cell comprising IL-1ra with (I) and causing IL-1ra to be
CC   released from the transformed cell, in a soluble form. The method is also
CC   useful for screening mutated cells or transformed cells producing
CC   elevated levels of intracellular POI. The method is used to release a POI
CC   for manufacturing food products, such as beverages, preparation of
CC   detergents, and in baking as a dough improving agent. The method obtains
CC   a fast, specific and economically efficient extraction of a soluble or
CC   membrane associated intracellular POI without the use of conventional
CC   cell disruption techniques. The resulting cell extract contains less
CC   contaminating intracellular DNA and is relatively free of cell wall
CC   fragments. The intracellular POI can be recovered from a eukaryotic host
CC   organism such as yeast, before glycosylation takes place. The method can
CC   be used to prevent contact of intracellular POI with the extracellular
CC   growth medium.
XX
SQ   Sequence 117 BP; 36 A; 23 C; 26 G; 32 T; 0 other;

  Query Match             7.1%;  Score 117;  DB 22;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-25;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1489 AAGAACGGTAAGTATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATC 1548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGAACGGTAAGTATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATC 60

Qy       1549 AAGGCCAAATGGTTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCT 1605
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGGCCAAATGGTTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCT 117


RESULT 15
AAS06162/c
ID   AAS06162 standard; DNA; 118 BP.
XX
AC   AAS06162;
XX
DT   12-SEP-2001  (first entry)
XX
DE   PCR primer hox10b used to clone synthetic HOX gene.
XX
KW   HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
KW   food manufacturing; beverage; detergent; baking; dough improving agent;
KW   D-hexose:O2-oxireductase; EC 1.1.3.5; PCR primer; hox10b; ss.
XX
OS   Chondrus crispus.
OS   Synthetic.
XX
PN   WO200138544-A1.
XX
PD   31-MAY-2001.
XX
PF   24-NOV-2000; 2000WO-IB01886.
XX
PR   24-NOV-1999;   99GB-0027801.
XX
PA   (DANI-) DANISCO AS.
XX
PI   Johansen CL,  Kjaerulff S,  Madrid SM,  Pedersen H,  Poulsen CH;
PI   Zargahi MR;
XX
DR   WPI; 2001-367695/38.
XX
PT   Releasing soluble or membrane associated intracellular protein from a
PT   cell for manufacturing food, comprises contacting the cell with a
PT   membrane extracting composition and causing the protein to be released
PT   in soluble form  -
XX
PS   Claim 19; Fig 5; 108pp; English.
XX
CC   The sequence represents the PCR primer hox10b used to clone synthetic
CC   hexose oxidase (D-hexose:O2-oxireductase, EC 1.1.3.5), (also referred to
CC   as HOX), gene. The native HOX gene was altered using site-directed
CC   mutagenesis in order to match the codon usage to known codon preferences
CC   of biotechnologically relevant yeasts, such as Pichia sp., to facilitate
CC   high level production in these organisms. The invention involves a method
CC   for releasing a soluble or membrane associated intracellular protein of
CC   interest (POI) from a cell involving contacting a cell comprising a
CC   soluble or membrane associated intracellular POI with a membrane
CC   extracting composition (I) and causing the POI to be released from the
CC   cell in a soluble form. The method is useful for releasing POI, such as
CC   an interleukin I receptor antagonist (IL-1ra) which involves contacting a
CC   transformed cell comprising IL-1ra with (I) and causing IL-1ra to be
CC   released from the transformed cell, in a soluble form. The method is also
CC   useful for screening mutated cells or transformed cells producing
```

CC elevated levels of intracellular POI. The method is used to release a POI
CC for manufacturing food products, such as beverages, preparation of
CC detergents, and in baking as a dough improving agent. The method obtains
CC a fast, specific and economically efficient extraction of a soluble or
CC membrane associated intracellular POI without the use of conventional
CC cell disruption techniques. The resulting cell extract contains less
CC contaminating intracellular DNA and is relatively free of cell wall
CC fragments. The intracellular POI can be recovered from a eukaryotic host
CC organism such as yeast, before glycosylation takes place. The method can
CC be used to prevent contact of intracellular POI with the extracellular
CC growth medium.
XX
SQ Sequence 118 BP; 39 A; 26 C; 20 G; 33 T; 0 other;

Query Match 7.1%; Score 116.4; DB 22; Length 118;
Best Local Similarity 99.2%; Pred. No. 6.5e-25;
Matches 117; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        732 GCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAGAATACTGTTGG 791
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        118 GCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAGAATACTGTTGG 59

Qy        792 TAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTGTATACATCC 849
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db         58 TAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTGTATACCTCC 1
```

Search completed: June 15, 2003, 02:56:53
Job time : 389 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2003, 02:45:32 ; Search time 4309 Seconds
(without alignments)
11103.507 Million cell updates/sec

Title: US-09-998-284-1
Perfect score: 1644
Sequence: 1 atggctactttgccacaaaa.....ctaagcagactaaatagtag 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1644	100.0	1644	6	AX155104	AX155104 Sequence
2		1638	99.6	1638	12	SCO294935	AJ294935 Synthetic
3		1280.8	77.9	1881	8	CCU89770	U89770 Chondrus cr
4		1279.2	77.8	1801	6	AR159722	AR159722 Sequence
5		1279.2	77.8	1801	6	AX030569	AX030569 Sequence
6	c	126	7.7	126	6	AX155099	AX155099 Sequence
7		120	7.3	120	6	AX155088	AX155088 Sequence
8	c	120	7.3	120	6	AX155097	AX155097 Sequence
9		118	7.2	118	6	AX155092	AX155092 Sequence
10		117	7.1	117	6	AX155102	AX155102 Sequence
11	c	116.4	7.1	118	6	AX155093	AX155093 Sequence
12	c	116	7.1	116	6	AX155091	AX155091 Sequence
13	c	111	6.8	111	6	AX155101	AX155101 Sequence
14	c	109	6.6	109	6	AX155089	AX155089 Sequence
15		109	6.6	109	6	AX155090	AX155090 Sequence
16		108	6.6	108	6	AX155098	AX155098 Sequence
17		108	6.6	108	6	AX155100	AX155100 Sequence
18	c	107	6.5	107	6	AX155085	AX155085 Sequence
19		106	6.4	106	6	AX155086	AX155086 Sequence
20	c	106	6.4	106	6	AX155087	AX155087 Sequence
21	c	102	6.2	102	6	AX155095	AX155095 Sequence
22		92.8	5.6	96	6	AX155094	AX155094 Sequence
23		85.4	5.2	10668	1	AE013658	AE013658 Yersinia
24		85.4	5.2	214050	1	AJ414142	AJ414142 Yersinia
25		85.2	5.2	90	6	AX155096	AX155096 Sequence
26		81.8	5.0	24225	1	SC2C4	AL512902 Streptomy
27	c	79.6	4.8	215050	1	AL646057	AL646057 Ralstonia
28		77.2	4.7	25883	1	AB008466	AB008466 Streptomy
29		69.2	4.2	26195	1	SC1G7	AL591083 Streptomy
30	c	63	3.8	78	6	AX155103	AX155103 Sequence
31		60	3.6	1620	6	AX089430	AX089430 Sequence
32		60	3.6	45624	6	AX089419	AX089419 Sequence
33	c	60	3.6	50000	6	AX089416	AX089416 Sequence
34	c	60	3.6	80161	1	AY007564	AY007564 Saccharop
35	c	60	3.6	80161	6	AR165018	AR165018 Sequence
36		56.8	3.5	2088	1	SPU80222	U80222 Streptomyce
37		56	3.4	61	6	AX155084	AX155084 Sequence
38		43.6	2.7	1141	6	AX083744	AX083744 Sequence
39		42.6	2.6	1738	8	AY051000	AY051000 Arabidops
40		42.6	2.6	89473	8	T1K7	AC013427 Sequence
41		42	2.6	1639	8	AY113892	AY113892 Arabidops
42		42	2.6	1785	8	AF360332	AF360332 Arabidops
43		42	2.6	51939	9	AC099341	AC099341 Homo sapi
44	c	42	2.6	91566	8	F25C20	AC007296 Arabidops
45		41.8	2.5	1584	6	AX412693	AX412693 Sequence

ALIGNMENTS

RESULT 1
AX155104
LOCUS AX155104 1644 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 22 from Patent WO0138544.
ACCESSION AX155104
VERSION AX155104.1 GI:14536652
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 1644)
AUTHORS Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H., Poulsen,C.H. and Zargahi,M.R.
TITLE Method for purifying proteins
JOURNAL Patent: WO 0138544-A 22 31-MAY-2001;

```
           DANISCO A/S (DK)
FEATURES             Location/Qualifiers
     source          1. .1644
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Nucleotide sequence"
BASE COUNT      430 a    337 c    404 g    473 t
ORIGIN

  Query Match           100.0%;  Score 1644;  DB 6;  Length 1644;
  Best Local Similarity 100.0%;  Pred. No. 0;
  Matches 1644;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240

Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300

Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260

Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380

Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500

Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620

Qy       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
              ||||||||||||||||||||||||
Db       1621 GAGCCTAAGCAGACTAAATAGTAG 1644


RESULT 2
SCO294935
LOCUS       SCO294935               1638 bp    DNA     linear   SYN 09-OCT-2000
DEFINITION  Synthetic construct partial sHOX gene for hexose oxidase.
ACCESSION   AJ294935
VERSION     AJ294935.1  GI:10798685
KEYWORDS    hexose oxidase; SHOX gene.
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1638)
  AUTHORS   Wolff,A., Hansen,O.C., Poulsen,U., Madrid,S. and Stougaard,P.
  TITLE     Recombinant production of hexose oxidase from the red alga Chondrus
            crispus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1638)
  AUTHORS   Stougaard,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2000) Stougarrd P., Enzyme Technology,
            Biotechnological Institute, Kogle Alle 2, DK-2970 Hoersholm,
```

DENMARK

```
FEATURES             Location/Qualifiers
     source          1. .1638
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Chondrus crispus sequence with codon usage
                     optimized to Pichia pastoris"
     gene            1. .1638
                     /gene="sHOX"
     CDS             1. .>1638
                     /gene="sHOX"
                     /EC_number="1.1.3.5"
                     /note="86% sequence homology to native gene; 100% identity
                     with native aa sequence"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="hexose oxidase"
                     /protein_id="CAC12830.1"
                     /db_xref="GI:10798686"
                     /translation="MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNI
                     DFVYVVYTPQGACTALDRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVES
                     GYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHG
                     LPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFK
                     DLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEF
                     VMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRH
                     TSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEV
                     PDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNL
                     KWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGALELYF
                     LGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLKEPKQTK"
BASE COUNT      428 a    337 c    402 g    471 t
ORIGIN
```

Query Match 99.6%; Score 1638; DB 12; Length 1638;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1638; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240

Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300

Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260

Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380

Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500

Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620

Qy       1621 GAGCCTAAGCAGACTAAA 1638
```

```
                  ||||||||||||||||||
Db       1621 GAGCCTAAGCAGACTAAA 1638


RESULT 3
CCU89770
LOCUS       CCU89770                1881 bp    mRNA    linear   PLN 25-OCT-2000
DEFINITION  Chondrus crispus hexose oxidase mRNA, complete cds.
ACCESSION   U89770
VERSION     U89770.1  GI:1877521
KEYWORDS    .
SOURCE      Chondrus crispus.
  ORGANISM  Chondrus crispus
            Eukaryota; Rhodophyta; Florideophyceae; Gigartinales;
            Gigartinaceae; Chondrus.
REFERENCE   1  (bases 1 to 1881)
  AUTHORS   Hansen,O.C. and Stougaard,P.
  TITLE     Hexose oxidase from the red alga Chondrus crispus. Purification,
            molecular cloning, and expression in Pichia pastoris
  JOURNAL   J. Biol. Chem. 272 (17), 11581-11587 (1997)
  MEDLINE   97269074
   PUBMED   9111074
REFERENCE   2  (bases 1 to 1881)
  AUTHORS   Stougaard,P. and Hansen,O.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1997) Biotechnological Institute, Koglevej 2,
            Hoersholm, DK-2970, Denmark
FEATURES             Location/Qualifiers
     source          1. .1881
                     /organism="Chondrus crispus"
                     /db_xref="taxon:2769"
     CDS             165. .1805
                     /codon_start=1
                     /product="hexose oxidase"
                     /protein_id="AAB49376.1"
                     /db_xref="GI:1877522"
                     /translation="MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNI
                     DFVYVVYTPQGACTALDRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVES
                     GYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHG
                     LPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHTGGGGNFGIITKYYFK
                     DLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEF
                     VMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRH
                     TSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEV
                     PDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNL
                     KWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGALELYF
                     LGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLKEPKQTK"
BASE COUNT      459 a    485 c    500 g    437 t
ORIGIN

  Query Match             77.9%;  Score 1280.8;  DB 8;  Length 1881;
  Best Local Similarity   86.2%;  Pred. No. 0;
  Matches 1417;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
              |||||||||  |  ||  ||  ||||| |||||  ||||| ||||||||  ||  ||
Db        165 ATGGCTACTCTTCCCCAGAAAGACCCCGGTTATATTGTAATTGATGTCAACGCGGGCACG 224

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
              ||  ||||||||  ||||||  |   |  ||  ||||||||||  ||  ||||||  |   |  ||||||
Db        225 CCGGACAAGCCGGACCCACGTCTCCCCTCCATGAAGCAGGGCTTCAACCGCCGCTGGATT 284

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
              ||  ||  ||  ||||||||||||||  |||||  ||||||||  ||||||||||||||||||  |
Db        285 GGAACTAATATCGATTTCGTTTATGTCGTGTACACTCCTCAAGGTGCTTGTACTGCACTT 344

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
              |||  |  ||||||||||||||||||||  |||||  |||||  |||||  ||  ||  ||
Db        345 GACCGTGCTATGGAAAAGTGTTCTCCCGGTACAGTCAGGATCGTCTCTGGCGGCCATTGC 404

Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
              |||||  ||||||||  ||  ||||||||  ||||||||  ||  ||||||||  ||||||  |  |||
Db        405 TACGAGGACTTCGTATTTGACGAATGCGTCAAGGCCATCATCAACGTCACTGGTCTCGTT 464
```

```
Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
              ||   ||||||  ||||||||||  ||||||||||||       ||  ||  ||  ||  |||||
Db        465 GAGAGTGGTTATGACGACGATAGGGGTTACTTCGTCAGCAGTGGAGATACAAATTGGGGC 524

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
              ||||||||||||||||||||||||||||||  ||||||  |  ||  ||  ||||||||  ||||||
Db        525 TCCTTCAAGACCTTGTTCAGAGACCACGGAAGAGTTCTTCCCGGGGGTTCCTGCTACTCC 584

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              |||||  |  |||||  ||||||||||  ||||||||||  |||||||||  |  |||||  ||
Db        585 GTCGGCCTCGGTGGCCACATTGTCGGCGGAGGTGACGGCATTTTGGCCCGCTTGCATGGC 644

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              |  ||  |||||||||  |   |||  ||  ||  ||  ||||||||||||||||  |  ||||||||||
Db        645 CTCCCCGTCGATTGGCTCAGCGGCGTGGAGGTCGTCGTTAAGCCAGTCCTCACCGAAGAC 704

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              ||  ||  ||  |||||  ||  |||||  ||||||||||  ||||||||  |||  |  ||||||||
Db        705 TCGGTACTCAAGTATGTGCACAAAGATTCCGAAGGCAACGACGGGGAGCTCTTTTGGGCA 764

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
              |||||  |||||  |||||  ||  |||||  ||  ||  ||||||||||||||||||||||||||
Db        765 CACACAGGTGGCGGTGGCGGAAACTTTGGAATCATCACCAAATACTACTTCAAGGATTTG 824

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
              ||  |||||||||  |  ||  ||||||||  ||  ||  |||||||||   ||||||||||||||
Db        825 CCCATGTCTCCACGGGGCGTCATCGCATCAAATTTACACTTCAGCTGGGACGGTTTCACG 884

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              ||||||||||||||  ||||||||||  ||||||||||  |  ||  ||||||||||||||
Db        885 AGAGATGCCTTGCAGGATTTGTTGACAAAGTACTTCAAACTTGCCAGATGTGATTGGAAG 944

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
              |||||  |||||  |||||  ||||||||||||  ||  |||||  ||||||||||  ||||||||||
Db        945 AATACGGTTGGCAAGTTTCAAATCTTCCATCAGGCAGCGGAAGAGTTTGTCATGTACTTG 1004

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900
              ||||||||||||||  ||||||||||  |  ||||||||||||||  |  ||||||||||||
Db       1005 TATACATCCTACTCGAACGACGCCGAGCGCGAAGTTGCCCAAGACCGTCACTATCATTTG 1064

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960
              ||||||||||  ||||||||||||||||||||||||||||  ||||||||  |||||  ||  |||
Db       1065 GAGGCTGACATAGAACAGATCTACAAAACATGCGAGCCCACCAAAGCGCTTGGCGGGCAT 1124

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020
              |||||  |||||  ||  |||||  ||  |  ||  |  |||||  ||||||||||||||  ||  |||
Db       1125 GCTGGGTGGGCGCCGTTCCCCGTGCGGCCGCGCAAGAGGCACACATCCAAGACGTCGTAT 1184

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080
              ||||||||||||||  ||||||||||  |||||||||  |  ||||||||  ||||||||  |||
Db       1185 ATGCATGACGAGACGATGGACTACCCCTTCTACGCGCTCACTGAGACGATCAACGGCTCC 1244

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140
              ||  ||  ||||||  |  ||  ||||||||||||||  ||||||||||||||  ||  ||  |||
Db       1245 GGGCCGAATCAGCGCGGCAAGTACAAGTCTGCGTACATGATCAAGGATTTCCCGGATTTC 1304

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200
              |||||  ||  ||  ||||||||||||||||  |||||  ||  |||||  ||||||||||||||||
Db       1305 CAGATCGACGTGATCTGGAAATACCTTACGGAGGTCCCGGACGGCTTGACTAGTGCCGAA 1364

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
              ||||||||||  |  ||  |||||  ||  |||||  ||||||||||||||||||  ||  |||
Db       1365 ATGAAGGATGCCTTACTCCAGGTGGACATGTTTGGTGGTGAGATTCACAAGGTGGTCTGG 1424

Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320
              |||||  ||  |||||  ||  |||  |  ||||||||||||||||||||||||||||||||
Db       1425 GATGCGACGGCAGTCGCGCAGCGCGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1484

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380
              ||||||||||||||||||||||  |||  |  ||||||||||||||||||||||||||||||
Db       1485 CAGGAAGAAGACAAGGATGCAGTGAACCTCAAGTGGATTAGAGACTTTTACGAGGAGATG 1544

Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440
```

```
Db       1545 TATGAGCCGTATGGCGGGGTTCCAGACCCCAACACGCAGGTGGAGAGTGGTAAAGGTGTG 1604

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500

Db       1605 TTTGAGGGATGCTACTTCAACTACCCGGATGTGGACTTGAACAACTGGAAGAACGGCAAG 1664

Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560

Db       1665 TATGGTGCCCTCGAACTTTACTTTTTGGGTAACCTGAACCGCCTCATCAAGGCCAAATGG 1724

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620

Db       1725 TTGTGGGATCCCAACGAGATCTTCACAAACAAACAGAGCATCCCTACTAAACCTCTTAAG 1784

Qy       1621 GAGCCTAAGCAGACTAAATAGTAG 1644

Db       1785 GAGCCCAAGCAGACGAAATAGTAG 1808
```

```
RESULT 4
AR159722
LOCUS       AR159722               1801 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 30 from patent US 6251626.
ACCESSION   AR159722
VERSION     AR159722.1  GI:16222490
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1801)
  AUTHORS   Stougaard,P. and Hansen,O.Cai.
  TITLE     Recombinant hexose oxidase, a method of producing same and use of
            such enzyme
  JOURNAL   Patent: US 6251626-A 30 26-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .1801
                     /organism="unknown"
BASE COUNT      441 a    448 c    492 g    420 t
ORIGIN
```

Query Match 77.8%; Score 1279.2; DB 6; Length 1801;
Best Local Similarity 86.1%; Pred. No. 0;
Matches 1416; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

```
Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60

Db         84 ATGGCTACTCTTCCTCAGAAAGACCCCGGTTATATTGTAATTGATGTCAACGCGGGCACC 143

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120

Db        144 GCGGACAAGCCGGACCCACGTCTCCCCTCCATGAAGCAGGGCTTCAACCGCCGCTGGATT 203

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180

Db        204 GGAACTAATATCGATTTCGTTTATGTCGTGTACACTCCTCAAGGTGCTTGTACTGCACTT 263

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240

Db        264 GACCGTGCTATGGAAAAGTGTTCTCCCGGTACAGTCAGGATCGTCTCTGGCGGCCATTGC 323

Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300

Db        324 TACGAGGACTTCGTATTTGACGAATGCGTCAAGGCCATCATCAACGTCACTGGTCTCGTT 383

Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360

Db        384 GAGAGTGGTTATGACGACGATAGGGGTTACTTCGTCAGCAGTGGAGATACAAATTGGGGC 443

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420

Db        444 TCCTTCAAGACCTTGTTCAGAGACCACGGAAGAGTTCTTCCCGGGGGTTCCTGCTACTCC 503

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480

Db        504 GTCGGCCTCGGTGGCCACATTGTCGGCGGAGGTGACGGCATTTTGGCCCGCTTGCATGGC 563

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540

Db        564 CTCCCCGTCGATTGGCTCAGCGGCGTGGAGGTCGTCGTTAAGCCAGTCCTCACCGAAGAC 623

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600

Db        624 TCGGTACTCAAGTATGTGCACAAAGATTCCGAAGGCAACGACGGGGAGCTCTTTTGGGCA 683

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660

Db        684 CACACAGGTGGCGGTGGCGGAAACTTTGGAATCATCACCAAATACTACTTCAAGGATTTG 743

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720

Db        744 CCCATGTCTCCACGGGGCGTCATCGCATCAAATTTACACTTCAGCTGGGACGGTTTCACG 803

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780

Db        804 AGAGATGCCTTGCAGGATTTGTTGACAAAGTACTTCAAACTTGCCAGATGTGATTGGAAG 863

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840

Db        864 AATACGGTTGGCAAGTTTCAAATCTTCCATCAGGCAGCGGAAGAGTTTGTCATGTACTTG 923

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900

Db        924 TATACATCCTACTCGAACGACGCCGAGCGCGAAGTTGCCCAAGACCGTCACTATCATTTG 983

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960

Db        984 GAGGCTGACATAGAACAGATCTACAAAACATGCGAGCCCACCAAAGCGCTTGGCGGGCAT 1043

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020

Db       1044 GCTGGGTGGGCGCCGTTCCCCGTGCGGCCGCGCAAGAGGCACACATCCAAGACGTCGTAT 1103

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080

Db       1104 ATGCATGACGAGACGATGGACTACCCCTTCTACGCGCTCACTGAGACGATCAACGGCTCC 1163

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140

Db       1164 GGGCCGAATCAGCGCGGCAAGTACAAGTCTGCGTACATGATCAAGGATTTCCCGGATTTC 1223

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200

Db       1224 CAGATCGACGTGATCTGGAAATACCTTACGGAGGTCCCGGACGGCTTGACTAGTGCCGAA 1283

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260

Db       1284 ATGAAGGATGCCTTACTCCAGGTGGACATGTTTGGTGGTGAGATTCACAAGGTGGTCTGG 1343

Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320

Db       1344 GATGCGACGGCAGTCGCGCAGCGCGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1403

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380

Db       1404 CAGGAAGAAGACAAGGATGCAGTGAACCTCAAGTGGATTAGAGACTTTTACGAGGAGATG 1463

Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440

Db       1464 TATGAGCCGTATGGCGGGGTTCCAGACCCCAACACGCAGGTGGAGAGTGGTAAAGGTGTG 1523

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500

Db       1524 TTTGAGGGATGCTACTTCAACTACCCGGATGTGGACTTGAACAACTGGAAGAACGGCAAG 1583

Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560
```

```
                ||||||||| | ||||||||||||||||||||||||||| |  | ||||||||||||||||
Db       1584 TATGGTGCCCTCGAACTTTACTTTTTGGGTAACCTGAACCGCCTCATCAAGGCCAAATGG 1643

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620
                ||||||||||| |||||||||||||||||||||||||  |||||||||||||||||||||
Db       1644 TTGTGGGATCCCAACGAGATCTTCACAAACAAACAGAGCATCCCTACTAAACCTCTTAAG 1703

Qy       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
                |||||  ||||||||  |||||||||
Db       1704 GAGCCCAAGCAGACGAAATAGTAG 1727


RESULT 5
AX030569
LOCUS       AX030569               1801 bp    DNA     linear   PAT 20-SEP-2000
DEFINITION  Sequence 30 from Patent EP1020523.
ACCESSION   AX030569
VERSION     AX030569.1  GI:10278111
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1801)
  AUTHORS   Hansen,O.C. and Stougaard,P.
  TITLE     Recombinant hexose oxidase, a method of producing same and use of
            such enzyme
  JOURNAL   Patent: EP 1020523-A 30 19-JUL-2000;
            BIOTEKNOLOGISK INST (DK)
FEATURES             Location/Qualifiers
     source          1. .1801
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     CDS             84. .1724
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC09603.1"
                     /db_xref="GI:10278112"
                     /translation="MATLPQKDPGYIVIDVNAGTADKPDPRLPSMKQGFNRRWIGTNI
                     DFVYVVYTPQGACTALDRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVES
                     GYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHG
                     LPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHTGGGGNFGIITKYYFK
                     DLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEF
                     VMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRH
                     TSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEV
                     PDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNL
                     KWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGALELYF
                     LGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLKEPKQTK"
BASE COUNT      441 a    448 c    492 g    420 t
ORIGIN

  Query Match             77.8%;  Score 1279.2;  DB 6;  Length 1801;
  Best Local Similarity   86.1%;  Pred. No. 0;
  Matches 1416;  Conservative    0;  Mismatches  228;  Indels    0;  Gaps    0;

Qy          1 ATGGCTACTTTGCCACAAAAGGACCCAGGTTACATTGTTATTGACGTCAACGCTGGTACT 60
              |||||||||  |  ||  ||  ||||| |||||  |||||  |||||||| ||  ||
Db         84 ATGGCTACTCTTCCTCAGAAAGACCCCGGTTATATTGTAATTGATGTCAACGCGGGCACC 143

Qy         61 CCAGACAAGCCTGACCCAAGATTGCCATCCATGAAGCAAGGTTTCAACAGAAGATGGATT 120
                |  ||||||||  ||||||  |   |  ||  ||||||||||| ||  ||||||  |   |  ||||||
Db        144 GCGGACAAGCCGGACCCACGTCTCCCCTCCATGAAGCAGGGCTTCAACCGCCGCTGGATT 203

Qy        121 GGTACCAACATCGATTTCGTTTACGTCGTTTACACTCCACAAGGTGCTTGTACTGCTTTG 180
              ||  ||  ||  ||||||||||||||  |||||  ||||||||  ||||||||||||||||  |
Db        204 GGAACTAATATCGATTTCGTTTATGTCGTGTACACTCCTCAAGGTGCTTGTACTGCACTT 263

Qy        181 GACAGAGCTATGGAAAAGTGTTCTCCAGGTACCGTCAGAATCGTTTCTGGTGGTCACTGT 240
              |||  |  ||||||||||||||||||||  |||||  |||||  |||||  ||  ||  ||
Db        264 GACCGTGCTATGGAAAAGTGTTCTCCCGGTACAGTCAGGATCGTCTCTGGCGGCCATTGC 323

Qy        241 TACGAAGACTTCGTTTTCGACGAATGTGTCAAGGCTATTATCAACGTTACTGGTTTGGTT 300
              |||||  ||||||||  ||  ||||||||  ||||||||  ||||||  |  |||
Db        324 TACGAGGACTTCGTATTTGACGAATGCGTCAAGGCCATCATCAACGTCACTGGTCTCGTT 383

Qy        301 GAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCTTCCGGTGACACCAACTGGGGT 360
              ||   ||||||  ||||||||||  ||||||||||||      ||  ||  ||  ||  |||||
Db        384 GAGAGTGGTTATGACGACGATAGGGGTTACTTCGTCAGCAGTGGAGATACAAATTGGGGC 443

Qy        361 TCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCC 420
              ||||||||||||||||||||||||||||||  ||||||  |  ||  ||  ||||||||  ||||||
Db        444 TCCTTCAAGACCTTGTTCAGAGACCACGGAAGAGTTCTTCCCGGGGGTTCCTGCTACTCC 503

Qy        421 GTCGGTTTGGGTGGTCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGT 480
              |||||  |  |||||  ||||||||||  ||||||||||  |||||||||  |  |||||  ||
Db        504 GTCGGCCTCGGTGGCCACATTGTCGGCGGAGGTGACGGCATTTTGGCCCGCTTGCATGGC 563

Qy        481 TTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAAGCCAGTCTTGACCGAAGAC 540
              |  ||  |||||||||  |   |||  ||  ||  ||  ||||||||||||||  |  |||||||||
Db        564 CTCCCCGTCGATTGGCTCAGCGGCGTGGAGGTCGTCGTTAAGCCAGTCCTCACCGAAGAC 623

Qy        541 TCTGTTCTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCT 600
              ||  ||  ||  |||||  ||  |||||  ||||||||||  ||||||||  |||  |  |||||||||
Db        624 TCGGTACTCAAGTATGTGCACAAAGATTCCGAAGGCAACGACGGGGAGCTCTTTTGGGCA 683

Qy        601 CACACTGGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTG 660
              |||||  |||||  |||||  ||  |||||  ||  ||  ||||||||||||||||||||||||||
Db        684 CACACAGGTGGCGGTGGCGGAAACTTTGGAATCATCACCAAATACTACTTCAAGGATTTG 743

Qy        661 CCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACACTTCTCTTGGGACGGTTTCACT 720
              ||  |||||||||  |  ||  ||||||||  ||  ||  |||||||||   ||||||||||||||
Db        744 CCCATGTCTCCACGGGGCGTCATCGCATCAAATTTACACTTCAGCTGGGACGGTTTCACG 803

Qy        721 AGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAG 780
              ||||||||||||||  ||||||||||  ||||||||||  |  ||  ||||||||||||||
Db        804 AGAGATGCCTTGCAGGATTTGTTGACAAAGTACTTCAAACTTGCCAGATGTGATTGGAAG 863

Qy        781 AATACTGTTGGTAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTG 840
              |||||  |||||  |||||  ||||||||||  ||  |||||  ||||||||||  |||||||||
Db        864 AATACGGTTGGCAAGTTTCAAATCTTCCATCAGGCAGCGGAAGAGTTTGTCATGTACTTG 923

Qy        841 TATACATCCTACTCTAACGACGCCGAGAGAGAAGTTGCCCAAGACAGACACTATCATTTG 900
              ||||||||||||||  ||||||||||  |  ||||||||||||||  |  ||||||||||||
Db        924 TATACATCCTACTCGAACGACGCCGAGCGCGAAGTTGCCCAAGACCGTCACTATCATTTG 983

Qy        901 GAGGCTGACATTGAACAGATCTACAAAACATGCGAGCCTACCAAAGCTCTTGGTGGTCAT 960
              ||||||||||  ||||||||||||||||||||||||||||  ||||||||  |||||  ||  |||
Db        984 GAGGCTGACATAGAACAGATCTACAAAACATGCGAGCCCACCAAAGCGCTTGGCGGGCAT 1043

Qy        961 GCTGGTTGGGCTCCTTTCCCTGTTAGACCTAGAAAGAGACACACATCCAAGACTTCTTAT 1020
              |||||  |||||  ||  |||||  ||  |  ||  |  |||||  ||||||||||||||  ||  |||
Db       1044 GCTGGGTGGGCGCCGTTCCCCGTGCGGCCGCGCAAGAGGCACACATCCAAGACGTCGTAT 1103

Qy       1021 ATGCATGACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCC 1080
              ||||||||||||||  ||||||||||  ||||||||  |  ||||||||  ||||||||  |||
Db       1104 ATGCATGACGAGACGATGGACTACCCCTTCTACGCGCTCACTGAGACGATCAACGGCTCC 1163

Qy       1081 GGTCCTAATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTC 1140
              ||  ||  ||||||  |  ||  ||||||||||||||  ||||||||||||||  ||  ||  |||
Db       1164 GGGCCGAATCAGCGCGGCAAGTACAAGTCTGCGTACATGATCAAGGATTTCCCGGATTTC 1223

Qy       1141 CAGATTGATGTTATCTGGAAATACCTTACTGAGGTTCCTGACGGTTTGACTAGTGCCGAA 1200
              |||||  ||  ||  ||||||||||||||||  |||||  ||  |||||  ||||||||||||||||
Db       1224 CAGATCGACGTGATCTGGAAATACCTTACGGAGGTCCCGGACGGCTTGACTAGTGCCGAA 1283

Qy       1201 ATGAAGGATGCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGG 1260
              ||||||||||  |  ||  |||||  ||  ||||||||||||||||||||||  ||  |||
Db       1284 ATGAAGGATGCCTTACTCCAGGTGGACATGTTTGGTGGTGAGATTCACAAGGTGGTCTGG 1343

Qy       1261 GATGCTACTGCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1320
              |||||  ||  |||||  ||  |||  |  ||||||||||||||||||||||||||||||||||
Db       1344 GATGCGACGGCAGTCGCGCAGCGCGAGTACATCATCAAACTGCAGTACCAGACATACTGG 1403

Qy       1321 CAGGAAGAAGACAAGGATGCAGTTAACTTGAAGTGGATTAGAGACTTTTACGAGGAGATG 1380
              ||||||||||||||||||||||  |||  |  ||||||||||||||||||||||||||||||
Db       1404 CAGGAAGAAGACAAGGATGCAGTGAACCTCAAGTGGATTAGAGACTTTTACGAGGAGATG 1463
```

```
Qy       1381 TATGAGCCTTATGGTGGTGTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTT 1440
             |||||||| ||||| || |||||||||| ||||| ||||| ||||||||||||||||||
Db       1464 TATGAGCCGTATGGCGGGGTTCCAGACCCCAACACGCAGGTGGAGAGTGGTAAAGGTGTG 1523

Qy       1441 TTTGAGGGATGCTACTTCAACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAG 1500
             |||||||||||||||||||||||||| ||||| |||||||||||||||||||||||| |||
Db       1524 TTTGAGGGATGCTACTTCAACTACCCGGATGTGGACTTGAACAACTGGAAGAACGGCAAG 1583

Qy       1501 TATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATCAAGGCCAAATGG 1560
             |||||||| | |||||||||||||||||||||||||| | | ||||||||||||||||
Db       1584 TATGGTGCCCTCGAACTTTACTTTTTGGGTAACCTGAACCGCCTCATCAAGGCCAAATGG 1643

Qy       1561 TTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCTACTAAACCTCTTAAG 1620
             |||||||||| |||||||||||||||||||||||||   ||||||||||||||||||||
Db       1644 TTGTGGGATCCCAACGAGATCTTCACAAACAAACAGAGCATCCCTACTAAACCTCTTAAG 1703

Qy       1621 GAGCCTAAGCAGACTAAATAGTAG 1644
             ||||| |||||||| |||||||||
Db       1704 GAGCCCAAGCAGACGAAATAGTAG 1727
```

RESULT 6
AX155099/c
LOCUS AX155099 126 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 17 from Patent WO0138544.
ACCESSION AX155099
VERSION AX155099.1 GI:14536647
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 126)
AUTHORS Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
Poulsen,C.H. and Zargahi,M.R.
TITLE Method for purifying proteins
JOURNAL Patent: WO 0138544-A 17 31-MAY-2001;
DANISCO A/S (DK)
FEATURES Location/Qualifiers
source 1. .126
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide"
BASE COUNT 33 a 35 c 23 g 35 t
ORIGIN

Query Match 7.7%; Score 126; DB 6; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1210 GCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGGGATGCTACT 1269
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GCTCTTCTTCAGGTTGATATGTTCGGTGGTGAGATTCACAAGGTTGTTTGGGATGCTACT 67

Qy       1270 GCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGGCAGGAAGAA 1329
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          66 GCAGTTGCTCAGAGAGAGTACATCATCAAACTGCAGTACCAGACATACTGGCAGGAAGAA 7

Qy       1330 GACAAG 1335
             ||||||
Db           6 GACAAG 1
```

RESULT 7
AX155088
LOCUS AX155088 120 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 6 from Patent WO0138544.
ACCESSION AX155088
VERSION AX155088.1 GI:14536636
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 120)
AUTHORS Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
Poulsen,C.H. and Zargahi,M.R.
TITLE Method for purifying proteins
JOURNAL Patent: WO 0138544-A 6 31-MAY-2001;
DANISCO A/S (DK)
FEATURES Location/Qualifiers
source 1. .120
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide"
BASE COUNT 26 a 26 c 31 g 37 t
ORIGIN

Query Match 7.3%; Score 120; DB 6; Length 120;
Best Local Similarity 100.0%; Pred. No. 5.1e-21;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        280 ATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCT 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATCAACGTTACTGGTTTGGTTGAATCTGGTTACGACGACGATAGAGGTTACTTCGTCTCT 60

Qy        340 TCCGGTGACACCAACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTG 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCGGTGACACCAACTGGGGTTCCTTCAAGACCTTGTTCAGAGACCACGGTAGAGTTTTG 120
```

RESULT 8
AX155097/c
LOCUS AX155097 120 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 15 from Patent WO0138544.
ACCESSION AX155097
VERSION AX155097.1 GI:14536645
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 120)
AUTHORS Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
Poulsen,C.H. and Zargahi,M.R.
TITLE Method for purifying proteins
JOURNAL Patent: WO 0138544-A 15 31-MAY-2001;
DANISCO A/S (DK)
FEATURES Location/Qualifiers
source 1. .120
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic oligonucleotide"
BASE COUNT 33 a 25 c 29 g 33 t
ORIGIN

Query Match 7.3%; Score 120; DB 6; Length 120;
Best Local Similarity 100.0%; Pred. No. 5.1e-21;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1027 GACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCCGGTCCT 1086
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 GACGAGACTATGGACTACCCTTTCTACGCTTTGACTGAGACTATCAACGGTTCCGGTCCT 61

Qy       1087 AATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTCCAGATT 1146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          60 AATCAGAGAGGTAAGTACAAGTCTGCTTACATGATCAAGGACTTTCCAGACTTCCAGATT 1
```

RESULT 9
AX155092
LOCUS AX155092 118 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 10 from Patent WO0138544.
ACCESSION AX155092
VERSION AX155092.1 GI:14536640
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.

```
REFERENCE   1  (bases 1 to 118)
  AUTHORS   Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
            Poulsen,C.H. and Zargahi,M.R.
  TITLE     Method for purifying proteins
  JOURNAL   Patent: WO 0138544-A 10 31-MAY-2001;
            DANISCO A/S (DK)
FEATURES             Location/Qualifiers
     source          1. .118
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Synthetic oligonucleotide"
BASE COUNT       30 a     27 c     22 g     39 t
ORIGIN

  Query Match             7.2%;  Score 118;  DB 6;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-20;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        639 CAAATACTACTTCAAGGATTTGCCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACA 698
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAAATACTACTTCAAGGATTTGCCAATGTCTCCAAGAGGTGTCATCGCTTCTAACTTACA 60

Qy        699 CTTCTCTTGGGACGGTTTCACTAGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTC 756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTTCTCTTGGGACGGTTTCACTAGAGATGCCTTGCAAGATTTGTTGACTAAGTACTTC 118


RESULT 10
AX155102
LOCUS       AX155102                 117 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 20 from Patent WO0138544.
ACCESSION   AX155102
VERSION     AX155102.1  GI:14536650
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 117)
  AUTHORS   Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
            Poulsen,C.H. and Zargahi,M.R.
  TITLE     Method for purifying proteins
  JOURNAL   Patent: WO 0138544-A 20 31-MAY-2001;
            DANISCO A/S (DK)
FEATURES             Location/Qualifiers
     source          1. .117
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Synthetic oligonucleotide"
BASE COUNT       36 a     23 c     26 g     32 t
ORIGIN

  Query Match             7.1%;  Score 117;  DB 6;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-20;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1489 AAGAACGGTAAGTATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATC 1548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGAACGGTAAGTATGGTGCCTTGGAACTTTACTTTTTGGGTAACCTGAACAGATTGATC 60

Qy       1549 AAGGCCAAATGGTTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCT 1605
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGGCCAAATGGTTGTGGGATCCTAACGAGATCTTCACAAACAAACAGTCTATCCCT 117


RESULT 11
AX155093/c
LOCUS       AX155093                 118 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 11 from Patent WO0138544.
ACCESSION   AX155093
VERSION     AX155093.1  GI:14536641
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 118)
  AUTHORS   Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
            Poulsen,C.H. and Zargahi,M.R.
  TITLE     Method for purifying proteins
  JOURNAL   Patent: WO 0138544-A 11 31-MAY-2001;
            DANISCO A/S (DK)
FEATURES             Location/Qualifiers
     source          1. .118
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Synthetic oligonucleotide"
BASE COUNT       39 a     26 c     20 g     33 t
ORIGIN

  Query Match             7.1%;  Score 116.4;  DB 6;  Length 118;
  Best Local Similarity   99.2%;  Pred. No. 4.8e-20;
  Matches  117;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        732 GCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAGAATACTGTTGG 791
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        118 GCAAGATTTGTTGACTAAGTACTTCAAGTTGGCTAGATGTGATTGGAAGAATACTGTTGG 59

Qy        792 TAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTGTATACATCC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db         58 TAAGTTCCAAATCTTCCACCAAGCAGCTGAAGAGTTTGTTATGTACTTGTATACCTCC 1


RESULT 12
AX155091/c
LOCUS       AX155091                 116 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 9 from Patent WO0138544.
ACCESSION   AX155091
VERSION     AX155091.1  GI:14536639
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 116)
  AUTHORS   Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
            Poulsen,C.H. and Zargahi,M.R.
  TITLE     Method for purifying proteins
  JOURNAL   Patent: WO 0138544-A 9 31-MAY-2001;
            DANISCO A/S (DK)
FEATURES             Location/Qualifiers
     source          1. .116
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Synthetic oligonucleotide"
BASE COUNT       33 a     31 c     22 g     30 t
ORIGIN

  Query Match             7.1%;  Score 116;  DB 6;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-20;
  Matches  116;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        547 CTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCTCACACT 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 CTTAAGTACGTTCACAAGGATTCCGAAGGTAACGACGGTGAGTTGTTTTGGGCTCACACT 57

Qy        607 GGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTGCC 662
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         56 GGTGGAGGTGGAGGTAACTTCGGTATTATCACCAAATACTACTTCAAGGATTTGCC 1


RESULT 13
AX155101/c
LOCUS       AX155101                 111 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 19 from Patent WO0138544.
ACCESSION   AX155101
VERSION     AX155101.1  GI:14536649
KEYWORDS    .
SOURCE      synthetic construct.
```

```
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 111)
  AUTHORS   Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
            Poulsen,C.H. and Zargahi,M.R.
  TITLE     Method for purifying proteins
  JOURNAL   Patent: WO 0138544-A 19 31-MAY-2001;
            DANISCO A/S (DK)
FEATURES             Location/Qualifiers
     source          1. .111
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Synthetic oligonucleotide"
BASE COUNT       30 a     30 c     21 g     30 t
ORIGIN

  Query Match             6.8%;  Score 111;  DB 6;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-18;
  Matches  111;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1399 GTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTTTTTGAGGGATGCTACTTC 1458
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        111 GTTCCAGACCCTAACACTCAGGTTGAGAGTGGTAAAGGTGTTTTTGAGGGATGCTACTTC 52

Qy       1459 AACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAGTATGGTGCC 1509
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         51 AACTACCCTGATGTTGACTTGAACAACTGGAAGAACGGTAAGTATGGTGCC 1


RESULT 14
AX155089/c
LOCUS       AX155089                 109 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 7 from Patent WO0138544.
ACCESSION   AX155089
VERSION     AX155089.1  GI:14536637
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 109)
  AUTHORS   Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
            Poulsen,C.H. and Zargahi,M.R.
  TITLE     Method for purifying proteins
  JOURNAL   Patent: WO 0138544-A 7 31-MAY-2001;
            DANISCO A/S (DK)
FEATURES             Location/Qualifiers
     source          1. .109
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Synthetic oligonucleotide"
BASE COUNT       35 a     38 c     20 g     16 t
ORIGIN

  Query Match             6.6%;  Score 109;  DB 6;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-18;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        375 GTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGG 434
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 GTTCAGAGACCACGGTAGAGTTTTGCCAGGTGGTTCCTGTTACTCCGTCGGTTTGGGTGG 50

Qy        435 TCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGTTTG 483
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db         49 TCACATTGTCGGTGGAGGTGACGGTATTTTGGCCAGATTGCACGGTTTG 1


RESULT 15
AX155090
LOCUS       AX155090                 109 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 8 from Patent WO0138544.
ACCESSION   AX155090
VERSION     AX155090.1  GI:14536638
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 109)
  AUTHORS   Johansen,C.L., kj Rulff,S.R., Madrid,S.M., Pedersen,H.,
            Poulsen,C.H. and Zargahi,M.R.
  TITLE     Method for purifying proteins
  JOURNAL   Patent: WO 0138544-A 8 31-MAY-2001;
            DANISCO A/S (DK)
FEATURES             Location/Qualifiers
     source          1. .109
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Synthetic oligonucleotide"
BASE COUNT       22 a     23 c     28 g     36 t
ORIGIN

  Query Match             6.6%;  Score 109;  DB 6;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-18;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        462 TTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAA 521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTTGGCCAGATTGCACGGTTTGCCAGTCGATTGGTTATCCGGTGTTGAAGTTGTCGTTAA 60

Qy        522 GCCAGTCTTGACCGAAGACTCTGTTCTTAAGTACGTTCACAAGGATTCC 570
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCAGTCTTGACCGAAGACTCTGTTCTTAAGTACGTTCACAAGGATTCC 109


Search completed: June 15, 2003, 04:51:33
Job time : 4312 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 10:00:02 ; Search time 25 Seconds
(without alignments)
2254.767 Million cell updates/sec

Title: US-09-998-284-2
Perfect score: 2997
Sequence: 1 MATLPQKDPGYIVIDVNAGT.....IFTNKQSIPTKPLKEPKQTK 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2997	100.0	546	10	US-09-998-284-2	Sequence 2, Appli
2	2989	99.7	546	10	US-09-824-053-31	Sequence 31, Appl
3	292.5	9.8	509	9	US-09-258-031B-75	Sequence 75, Appl
4	277	9.2	508	9	US-09-258-031B-73	Sequence 73, Appl
5	275	9.2	529	9	US-09-258-031B-16	Sequence 16, Appl
6	272	9.1	529	9	US-09-258-031B-20	Sequence 20, Appl
7	266	8.9	540	9	US-09-258-031B-58	Sequence 58, Appl
8	263	8.8	508	9	US-09-258-031B-71	Sequence 71, Appl
9	107	3.6	19	10	US-09-932-923-1	Sequence 1, Appli
10	96.5	3.2	254	9	US-09-880-748-1466	Sequence 1466, Ap
11	95	3.2	622	9	US-10-124-880-4	Sequence 4, Appli
12	95	3.2	657	10	US-09-815-242-12135	Sequence 12135, A
13	94	3.1	600	9	US-10-234-026-13	Sequence 13, Appl
14	94	3.1	657	10	US-09-815-242-13013	Sequence 13013, A
15	92	3.1	584	10	US-09-815-242-5013	Sequence 5013, Ap
16	92	3.1	589	10	US-09-815-242-10878	Sequence 10878, A
17	91.5	3.1	413	9	US-09-910-186A-22	Sequence 22, Appl
18	91.5	3.1	550	9	US-09-738-626-5417	Sequence 5417, Ap
19	91.5	3.1	852	9	US-10-011-588-25	Sequence 25, Appl
20	91	3.0	652	10	US-09-815-242-5896	Sequence 5896, Ap
21	90.5	3.0	715	10	US-09-841-132-329	Sequence 329, App
22	90	3.0	23	10	US-09-824-053-2	Sequence 2, Appli
23	90	3.0	23	10	US-09-824-053-9	Sequence 9, Appli
24	89.5	3.0	467	9	US-09-759-130B-176	Sequence 176, App
25	89.5	3.0	467	10	US-09-801-196-20	Sequence 20, Appl
26	89.5	3.0	511	9	US-10-011-366-20	Sequence 20, Appl
27	89.5	3.0	608	9	US-10-011-366-21	Sequence 21, Appl
28	89.5	3.0	2366	9	US-10-011-366-10	Sequence 10, Appl
29	89	3.0	639	10	US-09-782-906-2	Sequence 2, Appli
30	89	3.0	639	10	US-09-782-906-5	Sequence 5, Appli
31	88.5	3.0	255	9	US-09-880-748-1153	Sequence 1153, Ap
32	88.5	3.0	1016	12	US-10-007-693-95	Sequence 95, Appl
33	88.5	3.0	1206	12	US-10-032-717-2	Sequence 2, Appli
34	88	2.9	639	10	US-09-782-906-3	Sequence 3, Appli
35	88	2.9	639	10	US-09-782-906-4	Sequence 4, Appli
36	87	2.9	305	1	US-08-325-278-1	Sequence 1, Appli
37	87	2.9	434	1	US-08-325-278-3	Sequence 3, Appli
38	87	2.9	592	9	US-10-234-026-15	Sequence 15, Appl
39	87	2.9	3571	9	US-10-150-821-2	Sequence 2, Appli
40	87	2.9	3571	10	US-09-911-842-2	Sequence 2, Appli
41	86.5	2.9	412	9	US-09-829-378-2	Sequence 2, Appli
42	86.5	2.9	416	9	US-10-118-495-29	Sequence 29, Appl
43	86.5	2.9	3712	9	US-10-108-605-103	Sequence 103, App
44	86	2.9	253	9	US-09-880-748-1545	Sequence 1545, Ap
45	85.5	2.9	925	9	US-09-738-626-6444	Sequence 6444, Ap

ALIGNMENTS

RESULT 1
US-09-998-284-2
; Sequence 2, Application US/09998284
; Patent No. US20020106361A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, et al.
; TITLE OF INVENTION: COMPOSITION
; FILE REFERENCE: 674509-2035
; CURRENT APPLICATION NUMBER: US/09/998,284
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: PCT/IB00/00829
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: GB 9913050.2
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Chondrus crispus
US-09-998-284-2

Query Match 100.0%; Score 2997; DB 10; Length 546;
Best Local Similarity 100.0%; Pred. No. 3.6e-262;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60

Qy         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240
```

```
Qy        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300

Qy        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360

Qy        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420

Qy        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480

Qy        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540

Qy        541 EPKQTK 546
              ||||||
Db        541 EPKQTK 546
```

RESULT 2
US-09-824-053-31
; Sequence 31, Application US/09824053
; Patent No. US20020106725A1
; GENERAL INFORMATION:
; APPLICANT: Peter Stougaard
; Ole Cai Hansen
; TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A
; METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hunton & Williams
; STREET: 1900 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006-1109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/824,053
; FILING DATE: 03-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/669,304
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Stanislaus Aksman
; REGISTRATION NUMBER: 28,562
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 955-1926
; TELEFAX: (202) 778-2201
; TELEX: No. US20020106725A1e
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-824-053-31

Query Match 99.7%; Score 2989; DB 10; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.9e-261;

Matches 545; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db          1 MATLPQKDPGYIVIDVNAGTADKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60

Qy         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300

Qy        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360

Qy        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420

Qy        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480

Qy        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540

Qy        541 EPKQTK 546
              ||||||
Db        541 EPKQTK 546
```

RESULT 3
US-09-258-031B-75
; Sequence 75, Application US/09258031B
; Patent No. US20020168735A1
; GENERAL INFORMATION:
; APPLICANT: STUIVER, Maarten Hendrik
; APPLICANT: CUSTERS, Jerome Humbertina Henricus Victor
; APPLICANT: SELA-BURLAGE, Marianne Beatrix
; APPLICANT: MELCHERS, Leo Sjoerd
; APPLICANT: VAN DEVENTER-TROOST, Johanna Pieternella
; APPLICANT: LAGEWEG, Wessel
; APPLICANT: PONSTEIN, Anne Silene
; APPLICANT: LAGEWEG, Wessel
; APPLICANT: PONSTEIN, Anne Silene
; TITLE OF INVENTION: ANTIFUNGAL PROTEINS, DNA CODING
; TITLE OF INVENTION: THEREFOR, AND HOSTS INCORPORATING
; TITLE OF INVENTION: SAME.
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LADAS & PARRY
; STREET: 26 WEST 61 STREET
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10023 - 7604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.25" Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: WORDPERFECT 8

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/258,031B
;      FILING DATE:  25-FEB-1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP97/04923
;      FILING DATE:  04-SEP-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP97200831.2
;      FILING DATE:  19-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP96202466.7
;      FILING DATE:  04-SEP-1996
;  INFORMATION FOR SEQ ID NO:  75:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  509 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-258-031B-75

  Query Match             9.8%;  Score 292.5;  DB 9;  Length 509;
  Best Local Similarity   24.9%;  Pred. No. 6.4e-18;
  Matches  125;  Conservative   68;  Mismatches  197;  Indels  113;  Gaps   23;

Qy         73 RIVSGGHCYEDFVFDECVK---AIINVTGLVESGYD-DDRGYFVSSGDTNWGSFKTLFRD 128
              |: |||| :|   :   ::    :::::| :   | :    :|  | |  |       :
Db         82 RVRSGGHDFEALSYVSRIEKPFILLDLSKLKQINVDIESNSAWVQPGAT-LGELYYRIAE 140

Qy        129 HGRV--LPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYV 186
               ::   | | | |||:||:: ||| | | | :||  | :  |::|       |:  | :
Db        141 KSKIHGFPAGLCTSVGIGGYMTGGGYGTLMRKYGLAGDNVLDVKMV-------DANGKLL 193

Qy        187 HKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQD 246
               : : | |  ||||  |||| :|||:  :  | :|: |: |    :       |:   ||
Db        194 DRAAMGED--LFWAIRGGGGASFGIVLAWKIKLVPV-PKTVTVFTV-------TKTLEQD 243

Qy        247 LLTKYFKLARCDWKNTVGKF-QIFHQAAEEF-VMYLYTSYSNDAEREVAQDRHYHLEADI 304
              |          |: |: ||  :  ||  :  :  :  ||  : |          :
Db        244 ARLK----------TISKWQQISSKIIEEIHIRVVLRAAGNDGNKTVTMTYLGQFLGEK 292

Qy        305 EQIYKTCE---PTKAL----------------GGHAGWAPFPVRPRKRHTSKTSYMHDE 344
               : |  |   |   |                ||    :|  :  : :      |
Db        293 GTLLKVMEKAFPELGLTQKDCTEMSWIEAALFHGGFPTGSPIEILLQLKSPLGKDYF--- 349

Qy        345 TMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDA 404
                                 |  | :: :  |       |:| | |   | |:
Db        350 ---------------------KATSDFVKEPIPVIGFKGIFKRLIE---GNTT------ 378

Qy        405 LLQVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDK-DAVNLKWIRDFYEEMYE 462
               |    :|| : |:   |      |   : |: |   | | ||  :  : ||:: |  |
Db        379 FLNWTPYGGMMSKIPESAIPFPHRNGTLFKILYYANWLENDKTSSRKINWIKEIYNYM-A 437

Qy        463 PYGGVPDPNTQVESGKGVFEGCYFNYPDVDL-NNWKNGKYGALEL------YFLGNLNRL 515
              ||          | |      | || |:|   |  | |   :|       || || :||
Db        438 PY---------VSSNP---RQAYVNYRDLDFGQNKNNAKVNFIEAKIWGPKYFKGNFDRL 485

Qy        516 IKAKWLWDPNEIFTNKQSIPTKP 538
              :| |   ||   | ::|||| |
Db        486 VKIKTKVDPENFFRHEQSIPPMP 508


RESULT 4
US-09-258-031B-73
; Sequence 73, Application US/09258031B
; Patent No. US20020168735A1
;  GENERAL INFORMATION:
;    APPLICANT:  STUIVER, Maarten Hendrik
;    APPLICANT:  CUSTERS, Jerome Humbertina Henricus Victor
;    APPLICANT:  SELA-BURLAGE, Marianne Beatrix
;    APPLICANT:  MELCHERS, Leo Sjoerd
;    APPLICANT:  VAN DEVENTER-TROOST, Johanna Pieternella
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    TITLE OF INVENTION:  ANTIFUNGAL PROTEINS, DNA CODING
;    TITLE OF INVENTION:  THEREFOR, AND HOSTS INCORPORATING
;    TITLE OF INVENTION:  SAME.
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LADAS & PARRY
;      STREET:  26 WEST 61 STREET
;      CITY:  NEW YORK
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10023 - 7604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.25" Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS 95
;      SOFTWARE:  WORDPERFECT 8
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/258,031B
;      FILING DATE:  25-FEB-1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP97/04923
;      FILING DATE:  04-SEP-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP97200831.2
;      FILING DATE:  19-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP96202466.7
;      FILING DATE:  04-SEP-1996
;  INFORMATION FOR SEQ ID NO:  73:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  508 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-258-031B-73

  Query Match             9.2%;  Score 277;  DB 9;  Length 508;
  Best Local Similarity   23.6%;  Pred. No. 1.6e-16;
  Matches  120;  Conservative   67;  Mismatches  199;  Indels  122;  Gaps   20;

Qy         72 VRIVSGGHCYED----------FVFDECVKAIINV-----TGLVESGYDDDRGYFVSSG 115
              ::| |||| |:            || |      ::|     |  |::|       |:
Db         82 MKIRSGGHDYDGLSYVTYSGKPFFVLDMFNLRSVDVDVASKTAWVQTGAILGEVYYYI-- 139

Qy        116 DTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKP 175
                |   |||         | | | :||:|||| ||| | : | :|| ||      :|
Db        140 ---WEKSKTL------AYPAGICPTVGVGGHISGGGYGNMMRKYGLTVDNTIDARMV--- 187

Qy        176 VLTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFS 235
                |   | : :   | |  |:|| |||||::|::  |   :|   |  |    :  :
Db        188 ----DVNGKILDRKLMGED--LYWAINGGGGGSYGVVLAYKI-NLVEVPENVTVFRISRT 240

Qy        236 WDGFTRDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEEFVMYLYTSYSNDAEREVA 293
               :    | :           |:    |   ::| :   : |    :|         :|
Db        241 LEQNATDIIH-----------RWQQVAPKLPDELFIRTVIDVVNGTVSSQKTVRTTFIA 288

Qy        294 Q-----------DRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYM 341
                         :| :    ::  :   |  |  :      |     |   :
Db        289 MFLGDTTTLLSILNRRF---PELGLVRSDCTETSWIQSVLFWTNIQVGSSE--------- 336

Qy        342 HDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEM 401
                        |  :  : |    | || |: :      :: ||| :           |:
Db        337 -----------TLLLQRNQPVNYLKRKSDYVREPISRTGLESIWKKM---------IEL 375

Qy        402 KDALLQVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEM 460
              :   :  : :|||: ::    |    |   : |:||   |::|      ::  |  |: |
Db        376 EIPTMAFNPYGGEMGRISSTVTPFPYRAGNLWKIQYGANWRDETLTDRYMELTRKLYQFM 435

Qy        461 YEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLN-NWKNGK---------YGALELYFLG 510
```

```
                |:    :|              :||| ||||  |  |||         ||  : || |
Db        436 -TPFVS-KNPRQS----------FFNYRDVDLGINSHNGKISSYVEGKRYG--KKYFAG 480

Qy        511 NLNRLIKAKWLWDPNEIFTNKQSIPTKP 538
              |  ||:| |   |    | |:||||  |
Db        481 NFERLVKIKTRVDSGNFFRNEQSIPVLP 508


RESULT 5
US-09-258-031B-16
; Sequence 16, Application US/09258031B
; Patent No. US20020168735A1
;  GENERAL INFORMATION:
;    APPLICANT:  STUIVER, Maarten Hendrik
;    APPLICANT:  CUSTERS, Jerome Humbertina Henricus Victor
;    APPLICANT:  SELA-BURLAGE, Marianne Beatrix
;    APPLICANT:  MELCHERS, Leo Sjoerd
;    APPLICANT:  VAN DEVENTER-TROOST, Johanna Pieternella
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    TITLE OF INVENTION:  ANTIFUNGAL PROTEINS, DNA CODING
;    TITLE OF INVENTION:  THEREFOR, AND HOSTS INCORPORATING
;    TITLE OF INVENTION:  SAME.
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LADAS & PARRY
;      STREET:  26 WEST 61 STREET
;      CITY:  NEW YORK
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10023 - 7604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.25" Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS 95
;      SOFTWARE:  WORDPERFECT 8
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/258,031B
;      FILING DATE:  25-FEB-1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP97/04923
;      FILING DATE:  04-SEP-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP97200831.2
;      FILING DATE:  19-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP96202466.7
;      FILING DATE:  04-SEP-1996
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  529 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-258-031B-16

  Query Match             9.2%;  Score 275;  DB 9;  Length 529;
  Best Local Similarity   23.9%;  Pred. No. 2.6e-16;
  Matches  144;  Conservative   64;  Mismatches  197;  Indels  198;  Gaps   26;

Qy         22 DKPDP--------------RLPSMKQGF--NRRWIGTNIDFVYVVYTPQG--------AC 57
              |: ||               |:: | :  | |:  |     ::: | :          |
Db         34 DRADPSFPITGEVYTPGNSSFPTVLQNYIRNLRFNETTTPKPFLIITAEHVSHIQAAVVC 93

Qy         58 TALDRAMEKCSPGTVRIVSGGHCYED----------FVFDECVKAIINV-----TGLVE 101
                 :| : |          |||| ||            |: |      |||     |  |:
Db         94 GKQNRLLLKTR-------SGGHDYEGLSYLTNTNQPFFIVDMFNLRSINVDIEQETAWVQ 146

Qy        102 SGYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGL 161
              :|      |:  :  :|          ||   | | | :||:|||  ||| | | | :||
```

```
Db        147 AGATLGEVYYRIAEKSN---------KHG--FPAGVCPTVGVGGHFSGGGYGNLMRKYGL 195

Qy        162 PVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLP 221
               || :   :::       |   | : : | | |  |||| ||||| :||::  |  | |
Db        196 SVDNIVDAQII-------DVNGKLLDRKSMGED--LFWAITGGGGVSFGVVLAYKIK-LV 245

Qy        222 MSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLY 281
              |  |                       | : | : :|
Db        246 RVPEVVTV---------------------FTIERREEQN------------------- 263

Qy        282 TSYSNDAER--EVAQ--DRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSK 337
                 |  |||  :||   ||   |      |  |       ||    | ||
Db        264 --LSTIAERWVQVADKLDRDLFLRMTFSVINDT------NGGKTVRAIFP---------- 305

Qy        338 TSYMHDE-------TMDYPFYALTE-----------------------TINGSGPNQR- 365
              | |: :          |:|   | |                       |   |   ||
Db        306 TLYLGNSRNLVTLLNKDFPELGLQESDCTEMSWVESVLYYTGFPSGTPTTALLSRTPQRL 365

Qy        366 --GKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVWDAT 423
                 | || |:       | : |:: |          |::: :|  : :|| : ::   |
Db        366 NPFKIKSDYVQNPISKRQFEFIFERL---------KELENQMLAFNPYGGRMSEISEFAK 416

Qy        424 AVAQRE-YIIKLQYQTYWQEEDKDAVN--LKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
                  |   | |:||:  |::   :| |  | : |  |: |        |       |
Db        417 PFPHRSGNIAKIQYEVNWEDLSDEAENRYLNFTRLMYDYM--------TPFVSKNPRK-- 466

Qy        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLG-------NLNRLIKAKWLWDPNEIFTNKQS 533
                : || |:|:    :|:    |    |       |  ||:  |   ||:  | |:||
Db        467 ---AFLNYRDLDIGINSHGRNAYTEGMVYGHKYFKETNYKRLVSVKTKVDPDNFFRNEQS 523

Qy        534 IPT 536
              |||
Db        524 IPT 526


RESULT 6
US-09-258-031B-20
; Sequence 20, Application US/09258031B
; Patent No. US20020168735A1
;  GENERAL INFORMATION:
;    APPLICANT:  STUIVER, Maarten Hendrik
;    APPLICANT:  CUSTERS, Jerome Humbertina Henricus Victor
;    APPLICANT:  SELA-BURLAGE, Marianne Beatrix
;    APPLICANT:  MELCHERS, Leo Sjoerd
;    APPLICANT:  VAN DEVENTER-TROOST, Johanna Pieternella
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    TITLE OF INVENTION:  ANTIFUNGAL PROTEINS, DNA CODING
;    TITLE OF INVENTION:  THEREFOR, AND HOSTS INCORPORATING
;    TITLE OF INVENTION:  SAME.
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LADAS & PARRY
;      STREET:  26 WEST 61 STREET
;      CITY:  NEW YORK
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10023 - 7604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.25" Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS 95
;      SOFTWARE:  WORDPERFECT 8
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/258,031B
;      FILING DATE:  25-FEB-1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP97/04923
;      FILING DATE:  04-SEP-1997
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  EP97200831.2
;      FILING DATE:  19-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP96202466.7
;      FILING DATE:  04-SEP-1996
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  529 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-258-031B-20

  Query Match             9.1%;  Score 272;  DB 9;  Length 529;
  Best Local Similarity   23.5%;  Pred. No. 4.8e-16;
  Matches  142;  Conservative   65;  Mismatches  198;  Indels  198;  Gaps   25;

Qy         22 DKPDP--------------RLPSMKQGF--NRRWIGTNIDFVYVVYTPQG---------AC 57
              |: ||                |:: | :  | |:  |     ::: | :          |
Db         34 DRADPSFPITGEVYTPGNSSFPTVLQNYIRNLRFNETTTPKPFLIITAEHVSHIQAAVVC 93

Qy         58 TALDRAMEKCSPGTVRIVSGGHCYED-----------FVFDECVKAIINV-----TGLVE 101
                 :| : |          |||| ||             |: |      |||     |  |:
Db         94 GKQNRLLLKTR-------SGGHDYEGLSYLTNTNQPFFIVDMFNLRSINVDIEQETAWVQ 146

Qy        102 SGYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGL 161
              :|      |:  :  :|          ||   | | | :||:|||  ||| | | | :||
Db        147 AGATLGEVYYRIAEKSN---------KHG--FPAGVCPTVGVGGHFSGGGYGNLMRKYGL 195

Qy        162 PVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLP 221
               || :   :::       |   | : : | | |  |||| ||||| :||::  |  | |
Db        196 SVDNIVDAQII-------DVNGKLLDRKSMGED--LFWAITGGGGVSFGVVLAYKIK-LV 245

Qy        222 MSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLY 281
               | | '                        | : | : :|
Db        246 RVPEVVTV---------------------FTIERREEQN------------------- 263

Qy        282 TSYSNDAER--EVAQ--DRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSK 337
               |  |||  :||   ||   |       |  |       ||    | ||
Db        264 --LSTIAERWVQVADKLDRDLFLRMTFSVINDT------NGGKTVRAIFP---------- 305

Qy        338 TSYMHDE-------TMDYPFYALTE-----------------------TINGSGPNQR- 365
               | |: :           |:|   | |                       |   |   ||
Db        306 TLYLGNSRNLVTLLNKDFPELGLQESDCTEMSWVESVLYYTGFPSGTPTTALLSRTPQRL 365

Qy        366 --GKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVWDAT 423
                 | || |:        | : |:: :           |::: :|  : :|| : ::   |
Db        366 NPFKIKSDYVQNPISKRQFEFIFERM---------KELENQMLAFNPYGGRMSEISEFAK 416

Qy        424 AVAQRE-YIIKLQYQTYWQEEDKDAVN--LKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
                  |   | |:||:  |::   :| |  | : |  |: |       |
Db        417 PFPHRSGNIAKIQYEVNWEDLSDEAENRYLNFTRLMYDYMTPFVSKNP---------- 464

Qy        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLG-------NLNRLIKAKWLWDPNEIFTNKQS 533
               : || |:|:    :|:    |     |        |  ||:  |   ||:  | |:||
Db        465 -REAFLNYRDLDIGINSHGRNAYTEGMVYGHKYFKETNYKRLVSVKTKVDPDNFFRNEQS 523

Qy        534 IPT 536
               |||
Db        524 IPT 526


RESULT 7
US-09-258-031B-58
; Sequence 58, Application US/09258031B
; Patent No. US20020168735A1
;  GENERAL INFORMATION:
;    APPLICANT:  STUIVER, Maarten Hendrik
;    APPLICANT:  CUSTERS, Jerome Humbertina Henricus Victor
;    APPLICANT:  SELA-BURLAGE, Marianne Beatrix
;    APPLICANT:  MELCHERS, Leo Sjoerd
;    APPLICANT:  VAN DEVENTER-TROOST, Johanna Pieternella
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    TITLE OF INVENTION:  ANTIFUNGAL PROTEINS, DNA CODING
;    TITLE OF INVENTION:  THEREFOR, AND HOSTS INCORPORATING
;    TITLE OF INVENTION:  SAME.
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LADAS & PARRY
;      STREET:  26 WEST 61 STREET
;      CITY:  NEW YORK
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10023 - 7604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.25" Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS 95
;      SOFTWARE:  WORDPERFECT 8
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/258,031B
;      FILING DATE:  25-FEB-1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP97/04923
;      FILING DATE:  04-SEP-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP97200831.2
;      FILING DATE:  19-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP96202466.7
;      FILING DATE:  04-SEP-1996
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  540 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-258-031B-58

  Query Match             8.9%;  Score 266;  DB 9;  Length 540;
  Best Local Similarity   23.5%;  Pred. No. 1.7e-15;
  Matches  140;  Conservative   72;  Mismatches  187;  Indels  198;  Gaps   30;

Qy         20 TPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQG--ACTALD----RAMEKCSPGTVR 73
              |||  :   ||: | : |     |: |     ||:     |||     :|   |:  | |
Db         56 TPD--NSSFPSVLQAYIR-----NLRFNEST-TPKPILIITALHPSHIQAAVVCAK-THR 106

Qy         74 IV----SGGHCYED-----------FVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTN 118
              ::    |||| ||             || |      |||:       :|:  :  :
Db        107 LLMKTRSGGHDYEGLSYVTNSNQPFFVVDMFNLRSINVS------IEDETAWVQAGATLG 160

Qy        119 WGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLT 178
                 ::   : :    | | | :||:|||  ||| | |   :|| || :   :::
Db        161 EVYYRIAEKSNSHAFPAGVCPTVGVGGHFSGGGYGNLMGKYGLSVDNIVDAQLI------ 214

Qy        179 EDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDG 238
               |   | ::: | | |  |||| ||||| :||::  |  | |   |  |   |:
Db        215 -DVNGKLLNRKSMGED--LFWAITGGGGVSFGVVVAYKIK-LVRVPTTVTVFNVQ----- 265

Qy        239 FTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHY 298
               | : |:| |   :      |       |:  :   :  : :  :  |:   |
Db        266 --RTSEQNLSTIAHR-----W------IQVADKLDNDLFLRMTFNVINNTNGE------- 305

Qy        299 HLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETIN 358
                                | : |      ||          | |: : |      ||   :|
Db        306 ----------------KTIRG-----LFP----------TLYLGNST------ALVALLN 328

Qy        359 GSGPNQRGKYKSAYMIKDFPDFQIDV------IW----KYLTEVPDGL-TSA-------- 399
                              ||||:  :::       |     : |  | |  |:|
Db        329 ----------------KDFPELGVEISDCIEMSWIESVLFYTNFPIGTPTTALLSRTPQR 372

Qy        400 ----------------------------EMKDALLQVDMFGGEIHKVVWDATAVAQRE-Y 430
```

```
                                        |::: :|  : :|| : ::   |     |
Db        373 LNPFKIKSDYVKNTISKQGFESIFERMKELENQMLAFNPYGGRMSEISEFAKPFPHRSGN 432

Qy        431 IIKLQYQTYWQEEDKDAVN--LKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNY 488
              | |:||:  | |   :| |  | : |  |: |       |                : ||
Db        433 IAKIQYEVNWDELGVEAANRYLNFTRVMYDYMTPFVSKNP------------REAFLNY 479

Qy        489 PDVDLNNWKNGK--------YGALELYFL-GNLNRLIKAKWLWDPNEIFTNKQSIPT 536
                |:|:    :||        ||    ||   | ||   |   ||:  | |:|||||
Db        480 RDLDIGVNSHGKNAYGEGMVYG--HKYFKETNYKRLTMVKTRVDPSNFFRNEQSIPT 534


RESULT 8
US-09-258-031B-71
; Sequence 71, Application US/09258031B
; Patent No. US20020168735A1
;  GENERAL INFORMATION:
;    APPLICANT:  STUIVER, Maarten Hendrik
;    APPLICANT:  CUSTERS, Jerome Humbertina Henricus Victor
;    APPLICANT:  SELA-BURLAGE, Marianne Beatrix
;    APPLICANT:  MELCHERS, Leo Sjoerd
;    APPLICANT:  VAN DEVENTER-TROOST, Johanna Pieternella
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    APPLICANT:  LAGEWEG, Wessel
;    APPLICANT:  PONSTEIN, Anne Silene
;    TITLE OF INVENTION:  ANTIFUNGAL PROTEINS, DNA CODING
;    TITLE OF INVENTION:  THEREFOR, AND HOSTS INCORPORATING
;    TITLE OF INVENTION:  SAME.
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LADAS & PARRY
;      STREET:  26 WEST 61 STREET
;      CITY:  NEW YORK
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10023 - 7604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.25" Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS 95
;      SOFTWARE:  WORDPERFECT 8
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/258,031B
;      FILING DATE:  25-FEB-1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP97/04923
;      FILING DATE:  04-SEP-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP97200831.2
;      FILING DATE:  19-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP96202466.7
;      FILING DATE:  04-SEP-1996
;  INFORMATION FOR SEQ ID NO:  71:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  508 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-258-031B-71

  Query Match             8.8%;  Score 263;  DB 9;  Length 508;
  Best Local Similarity   23.2%;  Pred. No. 3e-15;
  Matches  118;  Conservative   67;  Mismatches  201;  Indels  122;  Gaps   20;

Qy         72 VRIVSGGHCYED----------FVFDECVKAIINV-----TGLVESGYDDDRGYFVSSG 115
              ::| |||| |:           || |      ::|     |  |::|      |:
Db         82 MKIRSGGHDYDGLSYVTYSGKPFFVLDMFNLRSVDVDVASKTAWVQTGAILGEVYYYI-- 139

Qy        116 DTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKP 175
               .     |   |||        | | | :||:|||| ||| | : | :|| ||      :|
```

```
Db        140 ---WEKSKTL------AYPAGICPTVGVGGHISGGGYGNMMRKYGLTVDNTIDARMV--- 187

Qy        176 VLTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFS 235
                  |   | : :   | |  |:||  |||||::|::  |   :|   | |    :  :
Db        188 ----DVNGKILDRKLMGED--LYWAINGGGGGSYGVVLAYKI-NLVEVPENVTVFRISRT 240

Qy        236 WDGFTRDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEEFVMYLYTSYSNDAEREVA 293
               :   | :           |:    |   ::| :   : |    :|         :|
Db        241 LEQNATDIIH-----------RWQQVAPKLPDELFIRTVIDVVNGTVSSQKTVRTTFIA 288

Qy        294 Q-----------DRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYM 341
                         :| :    ::  :   |  |  :     |    |   :
Db        289 MFLGDTTTLLSILNRRF---PELGLVRSDCTETSWIQSVLFWTNIQVGSSE--------- 336

Qy        342 HDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEM 401
                    |  :  : |    | || |: :      :: ||| :           |:
Db        337 -----------TLLLQRNQPVNYLKRKSDYVREPISRTGLESIWKKM---------IEL 375

Qy        402 KDALLQVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEM 460
              :   :  : :|||: ::    |    |   : |:||   |::|      ::  |  |: |
Db        376 EIPTMAFNPYGGEMGRISLRVTPFPYRAGNLWKIQYGANWRDETLTDRYMELTRKLYQFM 435

Qy        461 YEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLN-NWKNGK---------YGALELYFLG 510
               |:    :|           :||  ||||  |  |||          ||  : || |
Db        436 -TPFVS-KNPRQS----------FFNNRDVDLGINSHNGKISSYVEGKRYG--KKYFAG 480

Qy        511 NLNRLIKAKWLWDPNEIFTNKQSIPTKP 538
              |  ||:| |   |    | |: |||  |
Db        481 NFERLVKIKTRVDSGNFFRNEHSIPVLP 508


RESULT 9
US-09-932-923-1
; Sequence 1, Application US/09932923
; Patent No. US20020064577A1
; GENERAL INFORMATION:
;  APPLICANT: SOE, JORN B.
;  APPLICANT:  POULSEN, CHARLOTTE H.
;  APPLICANT:  HOSTRUP, PERNILLE B.
;  TITLE OF INVENTION: A METHOD OF IMPROVING THE PROPERTIES OF A FLOUR DOUGH,
;  TITLE OF INVENTION:  A FLOUR DOUGH IMPROVING COMPOSITION AND IMPROVED FOOD
;  TITLE OF INVENTION:  PRODUCTS
;  FILE REFERENCE: 59191-000003
;  CURRENT APPLICATION NUMBER: US/09/932,923
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: 08/676,186
;  PRIOR FILING DATE: 1996-07-12
;  PRIOR APPLICATION NUMBER: 08/483,870
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: PCT/DK96/00239
;  PRIOR FILING DATE: 1996-06-04
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 19
;   TYPE: PRT
;   ORGANISM: Chondrus Crispus
;   FEATURE:
;   OTHER INFORMATION: Description of Sequence:  peptide
US-09-932-923-1

  Query Match             3.6%;  Score 107;  DB 10;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 0.0033;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 DPGYIVIDVNAGTPDKPDP 26
              |||||||||||||||||||
Db          1 DPGYIVIDVNAGTPDKPDP 19


RESULT 10
US-09-880-748-1466
; Sequence 1466, Application US/09880748
```

; Publication No. US20030059937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind BLyS
; FILE REFERENCE: PF523
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1466
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-1466

Query Match 3.2%; Score 96.5; DB 9; Length 254;
Best Local Similarity 21.2%; Pred. No. 1.2;
Matches 44; Conservative 22; Mismatches 65; Indels 77; Gaps 6;

```
Qy          17 NAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTALDRAMEKCSPGTVRIVS  76
               |||  :| : |: ||     ::           ||    : |: | |:  |: ||
Db          57 NAGYAEKFEGRV-SMTTNIPKK----------TVYMELSSLTSEDTAVYYCARGT-----  100

Qy          77 GGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGG  136
                                  |||   ||:: |    ||            |
Db         101 -------------------------GYDILTGYYMGSAFDQWGR-------------G  120

Qy         137 SCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGE  196
               :  :|  ||   |||               ||     : |||:          : |  |:
Db         121 TLVTVSSGGGGSGGGG-------------SGGGGSAQSVLTQP-------PSASGTPGQ  159

Qy         197 LFWAHTGGGGGNFGIITKYYFKDLPMSP  224
                     |||  | |      ::: :| :|
Db         160 TVTISCSGGGTNIGRDRVTWYQQVPGTP  187
```

RESULT 11
US-10-124-880-4
; Sequence 4, Application US/10124880
; Publication No. US20030026810A1
; GENERAL INFORMATION:
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene No. US20030026810A1boe
; APPLICANT: Schulein, Martin
; APPLICANT: Outtrup, Helle
; TITLE OF INVENTION: No. US20030026810A1el Rhamnogalacturonan Hydrolases
; FILE REFERENCE: 5572.204-US
; CURRENT APPLICATION NUMBER: US/10/124,880
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US/09/311,626B
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: 0608/98
; PRIOR FILING DATE: 1998-05-01
; PRIOR APPLICATION NUMBER: 60/084,358
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-124-880-4

Query Match 3.2%; Score 95; DB 9; Length 622;
Best Local Similarity 20.4%; Pred. No. 5.9;
Matches 114; Conservative 81; Mismatches 212; Indels 152; Gaps 33;

```
Qy          31 MKQGFNRRW----IGTNIDFVYVV--YTPQ----GACTALDRAMEKCSPGTVRIVSGGHC  80
               ||:|  :||       :::  : :|  :: |    |   |  | ||  : | | : :|
Db           1 MKKG-KKRWKNLLAASSLLLITLVTGFSEQAEADGRTAAQARQMESLNRGLVAVKTGNGV  59

Qy          81 YEDF--VFDECVKAIINV----TGLVESGYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLP  134
               :  :  :  |      :||       |  |       |  : || |
Db          60 FVSWRLLGTEPSSVSLNVYRNGKKLNGSPITSSTNYQDAGGDLN---------------  103

Qy         135 GGSCYSVGLGGHIVGGGD-------GILARLH-GLPVDWLSGVEVVVKPVLTEDSVLKYV  186
                 : | |      :: | :        |:| : :  :|:   :| :          | | |  |
Db         104 --AVYQVRA---VLNGREQAPSESVGVLNKQYKSVPLQKPAGGK-------TPDGV-SYT  150

Qy         187 HKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQD  246
               :  :: : |:|       |   : |  |:        |    :::   | ||:| | | |
Db         151 YSANDASLGDLV------GDAQYEIFLKW-------DP----SNSKDNSQDGYTGDVLID  193

Qy         247 LLTKYFKL-ARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIE  305
                     :||      |:  :||      :|   :  :|   :  | : |:|
Db         194 A----YKLDGTMMWRINLGK---NIRAGAHYTQFLVYDFDGDGKAEIAM----------  235

Qy         306 QIYKTCEPTK-ALGGHAGWAPFPVR-PRKRHTSKTSYM---------HDETMDYP-----  349
                  || : ||   |   | |    |  : |  |   |:           |::|
Db         236 ---KTADGTKDGKGKVIGNANADYRNAQGRILSGPEYLTVFKGDTGAELTTVNYEPARGN  292

Qy         350 --------------FYALTETINGSGPN---QRGKYKSAYMIK-DFPDFQIDVIWKYLTE  391
                           | |    ::|  |:     || |    ::  :|   ::  :| : ::
Db         293 VADWGDSYGNRVDRFLAGVAYLDGERPSFVMARGYYTRTVLVAYNFRGGKLTKLWTFDSD  352

Qy         392 VP-DGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEED-KDAVN  449
                | :|  : :   :|   |: |    :::: | ||   ::  |  | | |   |     |
Db         353 APGNGAYAGQGNHSLSVADVDGDGKDEIIYGAMAV---DHDGKGLYSTGWGHGDAMHTGN  409

Qy         450 LKWIR------DFYEEMYEPYG-GVPDPNT-----QVESGKGVFEGCYFNYPDVD-----  492
                |  |        :|    |||    |  |      | :|| |  |      |:|
Db         410 LDPSRPGLEVFQVHENSNSPYGLSFRDAKTGKIIWGVHAGKDVGRGM---AADIDPRYEG  466

Qy         493 LNNWKNGKYGALELYFLGN  511
                  | ||      :   :||
Db         467 AEVWANGSLYTAKGVKIGN  485
```

RESULT 12
US-09-815-242-12135
; Sequence 12135, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

```
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12135
;   LENGTH: 657
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12135

  Query Match             3.2%;  Score 95;  DB 10;  Length 657;
  Best Local Similarity   19.1%;  Pred. No. 6.4;
  Matches   79;  Conservative   51;  Mismatches  139;  Indels  144;  Gaps   21;

Qy        209 FGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVG---- 264
              | | |  |:      | |    |||      |     |:: :| ::   | :  |
Db          6 FYITTPIYY------PSG----NLHIGHAYST--VAGDVIARYKRMQGYDVRYLTGTDEH 53

Qy        265 --KFQIFHQAA--------EEFV-----MYLYTSYSNDAEREVAQDRHYHLEADIEQIYK 309
                | |   | |        :| :     ::     |||      ::|| |:   :||:::
Db         54 GQKIQEKAQKAGKTEIEYLDEMIAGIKQLWAKLEISNDDFIRTTEERHKHV---VEQVFE 110

Qy        310 TC--EPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGK 367
                 :   || : ||   |         :| |   : :| | |   : | |  |:  :
Db        111 RLLKQGDIYLGEYEGWYSVP--------DETYYTESQLVD-PQYENGKIIGGKSPDSGHE 161

Qy        368 YKSAYMIKDFPDF-----QIDVIWKYLTEVPDGLTSAEMKDALLQ-------VDM----- 410
               :   ::|:   |        | : ::  : || :     |: ::          |:
Db        162 VE---LVKEESYFFNISKYTDRLLEFYDQNPDFIQPPSRKNEMINNFIKPGLADLAVSRT 218

Qy        411 -FGGEIH--------KVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFY---- 457
               |   :|          ||     |     ||  | | :    :|:   |   |  | :
Db        219 SFNWGVHVPSNPKHVVYVWIDALV---NYISALGYLS----DDESLFNKYWPADIHLMAK 271

Qy        458 ----------------------------------EEMYEPYGGVPDPNTQVE------- 475
                                                 :| :  | | |||  ::
Db        272 EIVRFHSIIWPILLMALDLPLPKKVFAHGWILMKDGKMSKSKGNVVDPNILIDRYGLDAT 331

Qy        476 ----------SGKGVF------EGCYFNYPDVDLNNWKNGKYGALELYFLGNL 512
                     ,    |||      |   |:  : || | |     :  || | |
Db        332 RYYLMRELPFGSDGVFTPEAFVERTNFDLAN-DLGNLVNRTISMINKYFDGEL 383


RESULT 13
US-10-234-026-13
; Sequence 13, Application US/10234026
; Publication No. US20030097679A1
; GENERAL INFORMATION:
;  APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnol
;  APPLICANT:  Guy Jerome Corneel Bauw
;  APPLICANT:  Mark William Davey
;  APPLICANT:  Jens Ostergaard
;  APPLICANT:  Marc Charles Ernest Van Montagu
;  TITLE OF INVENTION: PRODUCTION OF ASCORBIC ACID IN PLANTS
;  FILE REFERENCE: DECLE5.001C1
;  CURRENT APPLICATION NUMBER: US/10/234,026
;  CURRENT FILING DATE:  2000-08-29
;  PRIOR APPLICATION NUMBER: NL 1006000
;  PRIOR FILING DATE: 1997-05-07
;  PRIOR APPLICATION NUMBER: PCT/EP98/02830
;  PRIOR FILING DATE: 1998-05-07
;  PRIOR APPLICATION NUMBER: 09/423,468
;  PRIOR FILING DATE: 2000-02-15
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 600
;   TYPE: PRT
;   ORGANISM: Brassica oleracea
US-10-234-026-13

  Query Match             3.1%;  Score 94;  DB 9;  Length 600;
  Best Local Similarity   17.9%;  Pred. No. 6.9;
  Matches   75;  Conservative   55;  Mismatches  128;  Indels  160;  Gaps   18;

Qy        143 LGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHT 202
              :|| |  |  |  |||   |:|     :|:   ::|       : ||   || |||
Db        210 IGGIIQVGAHGTGARLP--PID----EQVIGMKLVTPAKGTIELSKD---NDPELFHLAR 260

Qy        203 GGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNT 262
               | ||  |:: :     |     |  :  : : |        |:::   : ||
Db        261 CGLGG-LGVVAEV---TLQCVERQELLEHTYVS-------TLEEIKKNHKKLL------- 302

Qy        263 VGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAG 322
                      :  : | |||  |::                    :  || |     |
Db        303 ---------STNKHVKYLYIPYTDTV-----------------VVVTCNPVSKWSG--- 332

Qy        323 WAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQI 382
                                                    |  : || :   :|   |
Db        333 ------------------------------------APKDKPKYTTEEALKHVRDLYR 354

Qy        383 DVIWKYLTE-----VPDG-------LTSAEMKDALLQVDMFG----GEIHKVVWDATAVA 426
              : | ||  :      ||        |:  |::| |: :|       |::::       |
Db        355 ESIVKYRVQDSSKKTPDSREPDINELSFTELRDKLIALDPLNDVHVGKVNQ--------A 406

Qy        427 QREYIIKLQ-YQTYWQEE--DKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEG 483
              : |:  | : |:  | :|     |    :|:      |   | | :  |: :
Db        407 EAEFWKKSEGYRVGWSDEILGFDCGGQQWV----SETCFPAGTLAKPSMK---------- 452

Qy        484 CYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEI---FTNKQSIPTKP 538
                                    :| ::  |  ||: : :  |: |   :| :   | |
Db        453 --------------------DLEYIEQLKELIQKEAIPAPSPIEQRWTGRSKSPMSP 489


RESULT 14
US-09-815-242-13013
; Sequence 13013, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13013
;   LENGTH: 657
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-13013
```

```
  Query Match             3.1%;  Score 94;  DB 10;  Length 657;
  Best Local Similarity   19.1%;  Pred. No. 7.9;
  Matches   79;  Conservative   51;  Mismatches  139;  Indels  144;  Gaps   21;

Qy        209 FGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVG---- 264
              | | |  |:      | |    |||       |     |:: :| ::   | :   |
Db          6 FYITTPIYY------PSG----NLHIGHAYST--VAGDVIARYKRMQGYDVRYLTGTDEH 53

Qy        265 --KFQIFHQAA--------EEFV-----MYLYTSYSNDAEREVAQDRHYHLEADIEQIYK 309
                | |   | |        :| :     ::     |||      ::|| |:   :||::: 
Db         54 GQKIQEKAQKAGKTEIEYLDEMIAGIKQLWAKLEISNDDFIRTTEERHKHV---VEQVFE 110

Qy        310 TC--EPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGK 367
                :    || : ||   |          :| |   : :| | |   : | | |  |:   :
Db        111 RLLKQGDIYLGEYEGWYSVP--------DETYYTESQLVD-PQYENGKIIGGKSPDSGHE 161

Qy        368 YKSAYMIKDFPDF-----QIDVIWKYLTEVPDGLTSAEMKDALLQ-------VDM----- 410
               :   ::|:   |       | : ::  : || :     |: ::         |:
Db        162 VE---LVKEESYFFNISKYTDRLLEFYDQNPDFIQPPSRKNEMINNFIKPGLADLAVSRT 218

Qy        411 -FGGEIH--------KVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFY---- 457
               |   :|          ||     |     ||  | | :    :|:   |  |  | :
Db        219 SFNWGVHVPSNPKHVVYVWIDALV---NYISALGYLS----DDESLFNKYWPADIHLMAK 271

Qy        458 ------------------------------------EEMYEPYGGVPDPNTQVE------- 475
                                                 :| :  | | |||  ::
Db        272 EIVRFHSIIWPILLMALDLPLPKKVFAHGWILMKDGKMSKSKGNVVDPNILIDRYGLDAT 331

Qy        476 ----------SGKGVF------EGCYFNYPDVDLNNWKNGKYGALELYFLGNL 512
                         |||       |    |:  : || |  |      :  || | |
Db        332 RYYLMRELPFGSDGVFTPEAFVERTNFDLAN-DLGNLVNRTISMVNKYFDGEL 383


RESULT 15
US-09-815-242-5013
; Sequence 5013, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5013
;   LENGTH: 584
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-09-815-242-5013

  Query Match             3.1%;  Score 92;  DB 10;  Length 584;
  Best Local Similarity   21.3%;  Pred. No. 10;
  Matches   77;  Conservative   38;  Mismatches  118;  Indels  128;  Gaps   18;

Qy        216 YFKDLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEE 275
              :|  || ||:  :   |      || |         |::: ||
Db        191 HFYALPQSPQ--LFKQL-LMGAGFDR---------YYQIVRC----------------- 220

Qy        276 FVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKT-----------CEPTKALGGHAGW 323
                :       | : |  |       |||  : |             |  |   |
Db        221 ---FRDEDLRGDRQPEFTQ-------IDIETTFLTPEEIQTYTENMLAEVMKETKGIEIS 270

Qy        324 APFP-------------VRPRKRHTSKTSYMHDETMDYPFYAL---------TETINGSG 361
              |||              :|  |   :  : |   |  |                : :|  |
Db        271 VPFPRMSYDEAMARYGSDKPDTRFAMELIDVADVVKDVDFKVFQAALENGGHVKALNAKG 330

Qy        362 PNQR---------GKYKSAYMIKDFPDFQID------VIWKYLTEVPDGL---TSAEMKD 403
                :          ||| | :  |     :::       | |:|||| | |   |:||: |
Db        331 AADKYSRKDMDNLGKYVSQFGAKGLAWLKVEEDGLKGPIAKFLTEVSDELIAATNAEVGD 390

Qy        404 ALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEE----DKDAVNLKWIRDF--- 456
              |   |||  :  ::|  |                  :|   :|    |:   |  || |: 
Db        391 IL----MFGADKPEIVAAALGAV----------RTRLGKELGLIDESKFNFLWIVDWPLF 436

Qy        457 -YEEMYEPYGGVPDPNTQVES---------GKGVFEGCYFNYPDVDLNNWKNGKYGALEL 506
               |:|    |      | || ::           |:   |    || || :: |  |:| : 
Db        437 EYDEEAGRYVSAHHPFTQPKAEDVARLATDPASVYAEAY----DVVLNGYELGG-GSLRI 491

Qy        507 Y 507
              :
Db        492 H 492


Search completed: June 10, 2003, 10:07:16
Job time : 27 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 09:58:02 ; Search time 18 Seconds
(without alignments)
892.495 Million cell updates/sec

Title: US-09-998-284-2
Perfect score: 2997
Sequence: 1 MATLPQKDPGYIVIDVNAGT.....IFTNKQSIPTKPLKEPKQTK 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2989	99.7	546	4	US-08-669-304-31	Sequence 31, Appl
2	635.5	21.2	539	4	US-09-036-987A-11	Sequence 11, Appl
3	635.5	21.2	539	4	US-09-370-700-11	Sequence 11, Appl
4	323	10.8	397	4	US-09-433-248A-2	Sequence 2, Appli
5	304	10.1	529	4	US-09-433-248A-6	Sequence 6, Appli
6	275	9.2	538	4	US-09-647-390-16	Sequence 16, Appl
7	266	8.9	540	4	US-09-647-390-18	Sequence 18, Appl
8	255	8.5	495	4	US-09-217-490-2	Sequence 2, Appli
9	117	3.9	129	4	US-09-433-248A-8	Sequence 8, Appli
10	107	3.6	19	4	US-08-676-186-1	Sequence 1, Appli
11	95	3.2	622	4	US-09-311-626B-4	Sequence 4, Appli
12	94	3.1	600	4	US-09-423-468A-13	Sequence 13, Appl
13	92	3.1	466	3	US-08-704-711A-17	Sequence 17, Appl
14	92	3.1	466	4	US-09-521-220-17	Sequence 17, Appl
15	92	3.1	944	4	US-08-652-877-12	Sequence 12, Appl
16	92	3.1	4655	4	US-08-652-877-84	Sequence 84, Appl
17	91.5	3.1	858	4	US-09-255-829-22	Sequence 22, Appl
18	91.5	3.1	858	4	US-09-255-829-29	Sequence 29, Appl
19	91.5	3.1	1169	4	US-09-255-829-20	Sequence 20, Appl
20	91	3.0	1861	2	US-08-790-912-4	Sequence 4, Appli
21	90.5	3.0	603	4	US-09-134-001C-5226	Sequence 5226, Ap
22	90.5	3.0	715	4	US-09-620-412C-329	Sequence 329, App
23	90	3.0	23	4	US-08-669-304-2	Sequence 2, Appli
24	90	3.0	23	4	US-08-669-304-9	Sequence 9, Appli
25	90	3.0	4655	4	US-08-652-877-90	Sequence 90, Appl
26	89.5	3.0	444	1	US-09-178-002-2	Sequence 2, Appli
27	89.5	3.0	467	1	US-09-178-002-4	Sequence 4, Appli
28	89.5	3.0	467	4	US-09-391-104-24	Sequence 24, Appl
29	89.5	3.0	468	4	US-08-448-489-13	Sequence 13, Appl
30	89.5	3.0	511	1	US-08-480-604A-20	Sequence 20, Appl
31	89.5	3.0	511	2	US-08-405-496A-20	Sequence 20, Appl
32	89.5	3.0	511	4	US-08-915-136-20	Sequence 20, Appl
33	89.5	3.0	511	4	US-08-957-310-20	Sequence 20, Appl
34	89.5	3.0	608	1	US-08-480-604A-21	Sequence 21, Appl
35	89.5	3.0	608	2	US-08-405-496A-21	Sequence 21, Appl
36	89.5	3.0	608	4	US-08-915-136-21	Sequence 21, Appl
37	89.5	3.0	608	4	US-08-957-310-21	Sequence 21, Appl
38	89.5	3.0	609	1	US-08-480-604A-30	Sequence 30, Appl
39	89.5	3.0	609	4	US-08-915-136-30	Sequence 30, Appl
40	89.5	3.0	2366	1	US-08-480-604A-10	Sequence 10, Appl
41	89.5	3.0	2366	2	US-08-405-496A-10	Sequence 10, Appl
42	89.5	3.0	2366	4	US-08-915-136-10	Sequence 10, Appl
43	89.5	3.0	2366	4	US-08-957-310-10	Sequence 10, Appl
44	89	3.0	751	4	US-08-969-415-2	Sequence 2, Appli
45	88	2.9	383	4	US-09-434-774-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-669-304-31
; Sequence 31, Application US/08669304
; Patent No. 6251626
; GENERAL INFORMATION:
; APPLICANT: Peter Stougaard
; APPLICANT: Ole Cai Hansen
; TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A
; TITLE OF INVENTION: METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hunton & Williams
; STREET: 1900 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006-1109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,304
; FILING DATE: 12 July 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,910
; FILING DATE: 7 June 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stanislaus Aksman
; REGISTRATION NUMBER: 28,562
; REFERENCE/DOCKET NUMBER: 320.000003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 955-1926
; TELEFAX: (202) 778-2201
; TELEX: No. 6251626e
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-669-304-31

Query Match 99.7%; Score 2989; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 3.1e-296;
Matches 545; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60

```
Db          1 MATLPQKDPGYIVIDVNAGTADKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60

Qy         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120
Db         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
Db        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240
Db        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300
Db        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300

Qy        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360
Db        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360

Qy        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420
Db        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420

Qy        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
Db        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480

Qy        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540
Db        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540

Qy        541 EPKQTK 546
Db        541 EPKQTK 546


RESULT 2
US-09-036-987A-11
; Sequence 11, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  539 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-11

  Query Match             21.2%;  Score 635.5;  DB 4;  Length 539;
  Best Local Similarity   30.2%;  Pred. No. 5e-56;
  Matches  174;  Conservative   91;  Mismatches  216;  Indels   95;  Gaps   16;

Qy          2 ATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTALD 61
Db         25 APMNRRTPGTEI------TVEPDDPRYPDLVVGHNPRFTG-KPERIHIASSAEDVVHAVA 77

Qy         62 RAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYF-VSSGDTNWG 120
Db         78 DAVR--TGRRVGVRSGGHCFENLVADPAIRVLVDLSELNRVYYDSTRGAFAIEAGAALGQ 135

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
Db        136 VYRTLFKNWGVTIPTGACPGVGAGGHILGGGYGPLSRRFGSVVDYLQGVEVVVVDQAGEV 195

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKD----------LPMSPRGVIA 229
Db        196 HIVE-ADRNSTGAGHDLWWAHTGGGGGNFGIVTRFWLRTPDVVSTDAAELLPRPPATVLL 254

Qy        230 SNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSY----- 284
Db        255 RSFHWPWHELTEQSFAVLLQNF----------GNWYEQHSAPESTQLGLFSTLVCAHR 302

Qy        285 -------------SNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRP 330
Db        303 QAGYVTLNVHLDGTDPNAERTLAE----HLSAINAQVGVT--PAEGLRETLPWL------ 350

Qy        331 RKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLT 390
Db        351 --RSTQVAG----------AIAE--GGEPGMQRTKVKAAYLRTGLSEAQLATVYRRLT 394

Qy        391 EVPDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNL 450
Db        395 -----VYGYDNPAAALLLLGYGGMANAVAPSATALAQRDSVLKALFVTNWSEPAEDERHL 449

Qy        451 KWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNN--WKNGKYGALELYF 508
Db        450 TWIRGFYREMYAETGGVPVPGTRV-------DGSYINYPDTDLADPLWNTSGVAWHDLYY 502

Qy        509 LGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLKEPKQ 544
Db        503 KDNYPRLQRAKARWDPQNIFQHGLSIKPPARLSPGQ 538


RESULT 3
US-09-370-700-11
; Sequence 11, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
```

; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-11

Query Match 21.2%; Score 635.5; DB 4; Length 539;
Best Local Similarity 30.2%; Pred. No. 5e-56;
Matches 174; Conservative 91; Mismatches 216; Indels 95; Gaps 16;

```
Qy          2 ATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTALD 61
              | : :: ||  :      | :  ||| | :  | | |: |   : :::  : :    |:
Db         25 APMNRRTPGTEI------TVEPDDPRYPDLVVGHNPRFTG-KPERIHIASSAEDVVHAVA 77

Qy         62 RAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYF-VSSGDTNWG 120
               |:   :   | : |||||:|: | |  :: ::::: |    ||  || | : :|
Db         78 DAVR--TGRRVGVRSGGHCFENLVADPAIRVLVDLSELNRVYYDSTRGAFAIEAGAALGQ 135

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ::|||:: |  :| |:|  || ||||:||| | |:|  |  ||:| ||||||   |
Db        136 VYRTLFKNWGVTIPTGACPGVGAGGHILGGGYGPLSRRFGSVVDYLQGVEVVVVDQAGEV 195

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKD----------LPMSPRGVIA 229
              :::   ::| |   :|:|||||||||||||:|::: :            ||  |  |:
Db        196 HIVE-ADRNSTGAGHDLWWAHTGGGGGNFGIVTRFWLRTPDVVSTDAAELLPRPPATVLL 254

Qy        230 SNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSY----- 284
              : |: |   |  :   ||  :            | :   | | |   : |:::
Db        255 RSFHWPWHELTEQSFAVLLQNF----------GNWYEQHSAPESTQLGLFSTLVCAHR 302

Qy        285 -------------SNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRP 330
                          :||| :|:    || |   |:  |  | : |     |
Db        303 QAGYVTLNVHLDGTDPNAERTLAE----HLSAINAQVGVT--PAEGLRETLPWL------ 350

Qy        331 RKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLT 390
              | |           |: |   |    || | |:||:     : |:  ::: ||
Db        351 --RSTQVAG----------AIAE--GGEPGMQRTKVKAAYLRTGLSEAQLATVYRRLT 394

Qy        391 EVPDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNL 450
                  :   :   | | :  :||  : |   |||:|||: ::|  : | | |  :|  :|
Db        395 -----VYGYDNPAAALLLLGYGGMANAVAPSATALAQRDSVLKALFVTNWSEPAEDERHL 449

Qy        451 KWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNN--WKNGKYGALELYF 508
              |||  ||  |||   |||| | |:|       :| | |||| || :  |        :||:
Db        450 TWIRGFYREMYAETGGVPVPGTRV-------DGSYINYPDTDLADPLWNTSGVAWHDLYY 502

Qy        509 LGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLKEPKQ 544
              |  || :||  |||  || :  ||       | |
Db        503 KDNYPRLQRAKARWDPQNIFQHGLSIKPPARLSPGQ 538
```

RESULT 4
US-09-433-248A-2
; Sequence 2, Application US/09433248A
; Patent No. 6355462
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Han, Feng
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: Disease Resistance Factors
; FILE REFERENCE: BB1252 US NA
; CURRENT APPLICATION NUMBER: US/09/433,248A
; CURRENT FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: 60/107,242
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Zea mays
US-09-433-248A-2

Query Match 10.8%; Score 323; DB 4; Length 397;
Best Local Similarity 27.9%; Pred. No. 2.4e-24;
Matches 120; Conservative 55; Mismatches 155; Indels 100; Gaps 21;

```
Qy        131 RVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHK 188
              |||  | | | :||:|||: ||| | | | :||  |       |:  ||  |:  : :::
Db         41 RVLAFPAGICPTVGVGGHLSGGGFGTLMRRYGLAAD------NVIDAVLV-DADGRLLNR 93

Qy        189 DSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQDLL 248
               : | |  ||||  |||| :||::  :  : |   | |     :  |        :  |:
Db         94 TTMGED--LFWAIRGGGGESFGVVLSWKLR-LVRVPETVTVFTVRRS----INQSASHLI 146

Qy        249 TKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIY 308
              ||:  :|     : : :  :  | |     ::|          | :   | : :  : |
Db        147 TKWQAIAPALPSDLILRVAVRSQHARFEALFL-----GRCSRLLEHMRVHFPDLGVTQ-- 199

Qy        309 KTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYA----LTETINGSGPNQ 364
              ||              |                |: : ||:    |   :: ||
Db        200 SDCE-------EISWI------------------QSTVYFAFYSSSKPLELLLDRSGETP 234

Qy        365 R-GKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVWDAT 423
              |  | || |: :  |    :  | :| | ||           || :| :|| :  :   ||
Db        235 RYVKAKSDYVQEPIPRHVWERTWSWL-EKPDA--------GLLILDPYGGRMGSISPSAT 285

Qy        424 AVAQRE-YIIKLQYQTYWQEEDKDAV--NLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
                  |:  :  ||| :|| |    |:   : |:|  |||| |||    :| |
Db        286 PFPHRKGNLYNLQYYSYWFENGTAALEKRMSWVRGLYEEM-EPYVS-KNPRTG------- 336

Qy        481 FEGCYFNYPDVDLNN--------------WKNGKYGALELYFLGNLNRLIKAKWLWDPN 525
                  | || |:||                |       | || ||  ||   | : |||
Db        337 ----YVNYRDLDLGTNELEDNVTSYARARIWG-------EKYFKGNFERLAAVKAMADPN 385

Qy        526 EIFTNKQSIP 535
              : | |:||||
Db        386 DFFRNEQSIP 395
```

RESULT 5
US-09-433-248A-6
; Sequence 6, Application US/09433248A
; Patent No. 6355462
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Han, Feng
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: Disease Resistance Factors
; FILE REFERENCE: BB1252 US NA
; CURRENT APPLICATION NUMBER: US/09/433,248A
; CURRENT FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: 60/107,242
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Glycine max
US-09-433-248A-6

Query Match 10.1%; Score 304; DB 4; Length 529;
Best Local Similarity 23.6%; Pred. No. 3.4e-22;
Matches 129; Conservative 77; Mismatches 205; Indels 136; Gaps 26;

```
Qy         36 NRRWIGTNIDFVYVVYTP------QGACTALDRAMEKCSPG---TVRIVSGGHCYEDFVF 86
              | ||:  :     :: ||      | |          ||      :|: |||| ||   :
```

```
Db        65 NPRWVNISSRKPLMILTPFHESEIQAAIL--------CSKELKLQLRVRSGGHDYEGLSY 116

Qy        87 DECVKAI----INVTGLVESGYDDDRGYF---VSSGDTNWGSFKTLFRDHGRVLPGGSCY 139

Db       117 LSDVPFVMVDLINIRS-IEINLADETAWVQAGASIGELYYKISKA-SKVHG--FPAGTCP 172

Qy       140 SVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFW 199

Db       173 SVGIGGHISGGGQGLMLRKHGLAADNVVDAYLI-------DANGKIHDRKSMGED--VFW 223

Qy       200 AHTGGGGGNFGIITKYYFK----------DLPMSPRGVIASNLHFSWDGFTRDALQDLL 248

Db       224 AIRGGDASSFGVILAWKIKLVRVPPIVTGFNVPRTPEEGVTDLIH-RWQYIAHDLHEDLV 282

Qy       249 TKYFKLARCDWKNTVGKFQ-----IFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEAD 303

Db       283 IRV--IAQISGHDKSKKFRATFNSIFLGGVDRLIPLMNESFP----------ELGLQA- 328

Qy       304 IEQIYKTC------EPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETI 357

Db       329 -----KDCTEMSWIQSVMFIAGYNIEDPLELLLNRTTMFKRSF---------------- 366

Qy       358 NGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHK 417

Db       367 ---------KAKSDFFKEPVPKSGLEGAWKLL---------LEEEIAFLIMEPYGGRMNE 408

Qy       418 VVWDATAVAQRE-YIIKLQYQTYWQEEDKDAV--NLKWIRDFYEEMYEPYGGVPDPNTQV 474

Db       409 ISESEIPFPHRKGNLYNLQYLVNWEVNSDEASRRHLQWAKMVYKYM-TPY---------V 458

Qy       475 ESGKGVFEGCYFNYPDVDL--NNWKNGKYGALEL----YFLGNLNRLIKAKWLWDPNEIF 528

Db       459 SKSP---RAAYFNYKDLDLGKNKLDSTSYSEASVWGKKYFKGNFRRLAQIKTKFDPLNFF 515

Qy       529 TNKQSIP 535

Db       516 RNEQSIP 522


RESULT 6
US-09-647-390-16
; Sequence 16, Application US/09647390
; Patent No. 6465636
; GENERAL INFORMATION:
;  APPLICANT:  Stuiver, Maarten
;  APPLICANT:  Custers, Jerome
;  APPLICANT:  Simons, Lambertus
;  TITLE OF INVENTION:  Pathogen-Inducible Promoter
;  FILE REFERENCE:  MOG 57707/UST
;  CURRENT APPLICATION NUMBER: US/09/647,390
;  CURRENT FILING DATE:  2000-09-29
;  PRIOR APPLICATION NUMBER:  EP 98201024.1
;  PRIOR FILING DATE:  1998-04-01
;  PRIOR APPLICATION NUMBER:  PCT/EP99/02178
;  PRIOR FILING DATE:  1999-03-25
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 16
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Helianthus annuus
US-09-647-390-16

  Query Match             9.2%;  Score 275;  DB 4;  Length 538;
  Best Local Similarity   23.9%;  Pred. No. 3.1e-19;
  Matches  144;  Conservative   64;  Mismatches  197;  Indels  198;  Gaps   26;

Qy        22 DKPDP--------------RLPSMKQGF--NRRWIGTNIDFVYVVYTPQG--------AC 57

Db        43 DRADPSFPITGEVYTPGNSSFPTVLQNYIRNLRFNETTTPKPFLIITAEHVSHIQAAVVC 102

Qy        58 TALDRAMEKCSPGTVRIVSGGHCYED----------FVFDECVKAIINV-----TGLVE 101


Db       516 RNEQSIP 522


RESULT 6
US-09-647-390-16
; Sequence 16, Application US/09647390
; Patent No. 6465636
; GENERAL INFORMATION:
;  APPLICANT:  Stuiver, Maarten
;  APPLICANT:  Custers, Jerome
;  APPLICANT:  Simons, Lambertus
;  TITLE OF INVENTION:  Pathogen-Inducible Promoter
;  FILE REFERENCE:  MOG 57707/UST
;  CURRENT APPLICATION NUMBER: US/09/647,390
;  CURRENT FILING DATE:  2000-09-29
;  PRIOR APPLICATION NUMBER:  EP 98201024.1
;  PRIOR FILING DATE:  1998-04-01
;  PRIOR APPLICATION NUMBER:  PCT/EP99/02178
;  PRIOR FILING DATE:  1999-03-25
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 16
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Helianthus annuus
US-09-647-390-16

  Query Match             9.2%;  Score 275;  DB 4;  Length 538;
  Best Local Similarity   23.9%;  Pred. No. 3.1e-19;
  Matches  144;  Conservative   64;  Mismatches  197;  Indels  198;  Gaps   26;

Qy        22 DKPDP--------------RLPSMKQGF--NRRWIGTNIDFVYVVYTPQG--------AC 57

Db        43 DRADPSFPITGEVYTPGNSSFPTVLQNYIRNLRFNETTTPKPFLIITAEHVSHIQAAVVC 102

Qy        58 TALDRAMEKCSPGTVRIVSGGHCYED----------FVFDECVKAIINV-----TGLVE 101


Db       103 GKQNRLLLKTR-------SGGHDYEGLSYLTNTNQPFFIVDMFNLRSINVDIEQETAWVQ 155

Qy       102 SGYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGL 161

Db       156 AGATLGEVYYRIAEKSN---------KHG--FPAGVCPTVGVGGHFSGGGYGNLMRKYGL 204

Qy       162 PVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLP 221

Db       205 SVDNIVDAQII-------DVNGKLLDRKSMGED--LFWAITGGGGVSFGVVLAYKIK-LV 254

Qy       222 MSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLY 281

Db       255 RVPEVVTV----------------------FTIERREEQN------------------- 272

Qy       282 TSYSNDAER--EVAQ--DRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSK 337

Db       273 --LSTIAERWVQVADKLDRDLFLRMTFSVINDT------NGGKTVRAIFP---------- 314

Qy       338 TSYMHDE-------TMDYPFYALTE-----------------------TINGSGPNQR- 365

Db       315 TLYLGNSRNLVTLLNKDFPELGLQESDCTEMSWVESVLYYTGFPSGTPTTALLSRTPQRL 374

Qy       366 --GKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVWDAT 423

Db       375 NPFKIKSDYVQNPISKRQFEFIFERL---------KELENQMLAFNPYGGRMSEISEFAK 425

Qy       424 AVAQRE-YIIKLQYQTYWQEEDKDAVN--LKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480

Db       426 PFPHRSGNIAKIQYEVNWEDLSDEAENRYLNFTRLMYDYM--------TPFVSKNPRK-- 475

Qy       481 FEGCYFNYPDVDLNNWKNGKYGALELYFLG-------NLNRLIKAKWLWDPNEIFTNKQS 533

Db       476 ---AFLNYRDLDIGINSHGRNAYTEGMVYGHKYFKETNYKRLVSVKTKVDPDNFFRNEQS 532

Qy       534 IPT 536

Db       533 IPT 535


RESULT 7
US-09-647-390-18
; Sequence 18, Application US/09647390
; Patent No. 6465636
; GENERAL INFORMATION:
;  APPLICANT:  Stuiver, Maarten
;  APPLICANT:  Custers, Jerome
;  APPLICANT:  Simons, Lambertus
;  TITLE OF INVENTION:  Pathogen-Inducible Promoter
;  FILE REFERENCE:  MOG 57707/UST
;  CURRENT APPLICATION NUMBER: US/09/647,390
;  CURRENT FILING DATE:  2000-09-29
;  PRIOR APPLICATION NUMBER:  EP 98201024.1
;  PRIOR FILING DATE:  1998-04-01
;  PRIOR APPLICATION NUMBER:  PCT/EP99/02178
;  PRIOR FILING DATE:  1999-03-25
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 540
;   TYPE: PRT
;   ORGANISM: Lactuca sativa
US-09-647-390-18

  Query Match             8.9%;  Score 266;  DB 4;  Length 540;
  Best Local Similarity   23.5%;  Pred. No. 2.6e-18;
  Matches  140;  Conservative   72;  Mismatches  187;  Indels  198;  Gaps   30;

Qy        20 TPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQG--ACTALD----RAMEKCSPGTVR 73

Db        56 TPD--NSSFPSVLQAYIR-----NLRFNEST-TPKPILIITALHPSHIQAAVVCAK-THR 106

Qy        74 IV----SGGHCYED----------FVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTN 118
```

```
Db        107 LLMKTRSGGHDYEGLSYVTNSNQPFFVVDMFNLRSINVS------IEDETAWVQAGATLG 160

Qy        119 WGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLT 178

Db        161 EVYYRIAEKSNSHAFPAGVCPTVGVGGHFSGGGYGNLMGKYGLSVDNIVDAQLI------ 214

Qy        179 EDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDG 238

Db        215 -DVNGKLLNRKSMGED--LFWAITGGGGVSFGVVVAYKIK-LVRVPTTVTVFNVQ----- 265

Qy        239 FTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHY 298

Db        266 --RTSEQNLSTIAHR-----W------IQVADKLDNDLFLRMTFNVINNTNGE------- 305

Qy        299 HLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETIN 358

Db        306 ----------------KTIRG-----LFP----------TLYLGNST------ALVALLN 328

Qy        359 GSGPNQRGKYKSAYMIKDFPDFQIDV------IW----KYLTEVPDGL-TSA-------- 399

Db        329 ----------------KDFPELGVEISDCIEMSWIESVLFYTNFPIGTPTTALLSRTPQR 372

Qy        400 --------------------------EMKDALLQVDMFGGEIHKVVWDATAVAQRE-Y 430

Db        373 LNPFKIKSDYVKNTISKQGFESIFERMKELENQMLAFNPYGGRMSEISEFAKPFPHRSGN 432

Qy        431 IIKLQYQTYWQEEDKDAVN--LKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNY 488

Db        433 IAKIQYEVNWDELGVEAANRYLNFTRVMYDYMTPFVSKNP------------REAFLNY 479

Qy        489 PDVDLNNWKNGK--------YGALELYFL-GNLNRLIKAKWLWDPNEIFTNKQSIPT 536

Db        480 RDLDIGVNSHGKNAYGEGMVYG--HKYFKETNYKRLTMVKTRVDPSNFFRNEQSIPT 534
```

RESULT 8
US-09-217-490-2
; Sequence 2, Application US/09217490
; Patent No. 6165761
; GENERAL INFORMATION:
; APPLICANT: Schneider, Palle
; APPLICANT: Christensen, Soren
; APPLICANT: Dybdal, Lone
; APPLICANT: Fuglsang, Claus Crone
; APPLICANT: Xu, Feng
; APPLICANT: Golightly, Elizabeth
; TITLE OF INVENTION: Carbohydrate Oxidase And use Thereof In
; TITLE OF INVENTION: Baking
; FILE REFERENCE: 5421.200-US
; CURRENT APPLICATION NUMBER: US/09/217,490
; CURRENT FILING DATE: 1998-12-21
; EARLIER APPLICATION NUMBER: PA 1997 01505
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: PA 1998 00763
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/068,717
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 60/088/725
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Microdochium nivale
US-09-217-490-2

Query Match 8.5%; Score 255; DB 4; Length 495;
Best Local Similarity 24.2%; Pred. No. 3e-17;
Matches 121; Conservative 77; Mismatches 175; Indels 126; Gaps 27;

```
Qy         76 SGGHCYEDFVF-DECVKAIINVTGLVE-SGYDDDRGYFVSSGDTNWGSFKTLFRD-HGRV 132
```

```
Db         89 SGGHSYASFGFGGENGHLMVQLDRMIDVISYNDKTGIAHVEPGARLGHLATVLNDKYGRA 148

Qy        133 LPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEG 192

Db        149 ISHGTCPGVGISGHFAHGGFGFSSHMHGLAVDSVVGVTV----VLADGRIV----EASAT 200

Qy        193 NDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYF 252

Db        201 ENADLFWG-IKGAGSNFGIVAVWKLATFP-APK---------------------VLTRFG 237

Qy        253 KLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL------EADIEQ 306

Db        238 --VTLNWKNKTSALKGI-EAVEDYARWV-------APREV----NFRIGDYGAGNPGIEG 283

Qy        307 I-YKTCEPTKALGG------HAGWAPFPVRPRKRHTSKTSYMHDETMDY-------PFYA 352

Db        284 LYYGTPEQWRAAFQPLLDTLPAGYVVNPTTSLNWIESVLSYSNFDHVDFITPQPVENFYA 343

Qy        353 LTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDA--LLQVDM 410

Db        344 KSLTL---------KSIKGDAVKNFVDYYFDV------------SNKVKDRFWFYQLDV 381

Qy        411 FGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDP 470

Db        382 HGGKNSQV----TKVTNAE--------TAYPHRDK-----LWLIQFYDR-YDNNQTYPET 423

Qy        471 NTQVESG------KGVFE---GCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWL 521

Db        424 SFKFLDGWVNSVTKALPKSDWGMYINYADPRMD-----RDYATKVYYGENLARLQKLKAK 478

Qy        522 WDPNEIFTNKQSIPTKPLK 540

Db        479 FDPTDRFYYPQAV--RPVK 495
```

RESULT 9
US-09-433-248A-8
; Sequence 8, Application US/09433248A
; Patent No. 6355462
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Han, Feng
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: Disease Resistance Factors
; FILE REFERENCE: BB1252 US NA
; CURRENT APPLICATION NUMBER: US/09/433,248A
; CURRENT FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: 60/107,242
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-433-248A-8

Query Match 3.9%; Score 117; DB 4; Length 129;
Best Local Similarity 27.8%; Pred. No. 0.00042;
Matches 40; Conservative 17; Mismatches 57; Indels 30; Gaps 6;

```
Qy        405 LLQVDMFGGEIHKVVWDATAVAQRE---YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMY 461

Db          1 LIVLEPHGGLMATIPTAATPYPHRSGVLYII--QYIAFWQGDGGTAAT-TWLGSFYDFMG 57

Qy        462 EPYGGVPDPNTQVESGKGVFEGCYFNYPDVDL----------NNWKNGKYGALELYFLGN 511

Db         58 HYVSSNP------------RQAYVNFRDLDIGQNAVSDDLSTTSQSGKVWG-ERYFMSN 103

Qy        512 LNRLIKAKWLWDPNEIFTNKQSIP 535

Db        104 YQRLASVKAAVDPMDYFRNEQSIP 127
```

```
RESULT 10
US-08-676-186-1
; Sequence 1, Application US/08676186
; Patent No. 6358543
;  GENERAL INFORMATION:
;       APPLICANT: SOE, Jorn B.
;                  POULSEN, Charlotte H.
;                  HOSTRUP, Pernille B.
;       TITLE OF INVENTION: A method of improving the
;                           properties of a flour dough, a flour dough improving
;                           composition and improved food products
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: McDermott, Will & Emery
;            STREET: 1850 K Street N.W., Suite 450
;            CITY: Washington
;            STATE: DC
;            COUNTRY: USA
;            ZIP: 20006-2296
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: Windows 95
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/676,186
;            FILING DATE: 12-Sep-1996
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/483,870; PCT/DK96/00239
;            FILING DATE: 07-JUN-95; 04-JUN-96
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Aksman, Stanislaus
;            REGISTRATION NUMBER: 28,562
;            REFERENCE/DOCKET NUMBER: 44648/013
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 202-778-8300
;            TELEFAX: 202-778-8335
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 19 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-676-186-1

  Query Match             3.6%;  Score 107;  DB 4;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 0.0002;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 DPGYIVIDVNAGTPDKPDP 26
              |||||||||||||||||||
Db          1 DPGYIVIDVNAGTPDKPDP 19


RESULT 11
US-09-311-626B-4
; Sequence 4, Application US/09311626B
; Patent No. 6399347
; GENERAL INFORMATION:
;  APPLICANT: Jorgensen, Per Lina
;  APPLICANT:  Schnorr, Kirk
;  APPLICANT:  Andersen, Lene No. 6399347boe
;  APPLICANT:  Schulein, Martin
;  APPLICANT:  Outtrup, Helle
;  TITLE OF INVENTION: No. 6399347el Rhamnogalacturonan Hydrolases
;  FILE REFERENCE: 5572.204-US
;  CURRENT APPLICATION NUMBER: US/09/311,626B
;  CURRENT FILING DATE:  1999-05-13
;  PRIOR APPLICATION NUMBER: 0608/98
;  PRIOR FILING DATE: 1998-05-01
;  PRIOR APPLICATION NUMBER: 60/084,358
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 622
;   TYPE: PRT
;   ORGANISM: Bacillus licheniformis
US-09-311-626B-4

  Query Match             3.2%;  Score 95;  DB 4;  Length 622;
  Best Local Similarity   20.4%;  Pred. No. 0.94;
  Matches  114;  Conservative   81;  Mismatches  212;  Indels  152;  Gaps   33;

Qy         31 MKQGFNRRW----IGTNIDFVYVV--YTPQ----GACTALDRAMEKCSPGTVRIVSGGHC 80
              ||:|  :||       :::  : :|  :: |     |   |  | ||  : | | : :|
Db          1 MKKG-KKRWKNLLAASSLLLITLVTGFSEQAEADGRTAAQARQMESLNRGLVAVKTGNGV 59

Qy         81 YEDF--VFDECVKAIINV----TGLVESGYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLP 134
              :  :  :  |      :||       |  |        |  : || |
Db         60 FVSWRLLGTEPSSVSLNVYRNGKKLNGSPITSSTNYQDAGGDLN---------------- 103

Qy        135 GGSCYSVGLGGHIVGGGD-------GILARLH-GLPVDWLSGVEVVVKPVLTEDSVLKYV 186
               : | |     :: | :        |:| : :  :|:   :| :       | | |  |
Db        104 --AVYQVRA---VLNGREQAPSESVGVLNKQYKSVPLQKPAGGK-------TPDGV-SYT 150

Qy        187 HKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQD 246
              :  :: : |:|        |   : :|  |:         |    :::   | ||:| | | |
Db        151 YSANDASLGDLV------GDAQYEIFLKW-------DP----SNSKDNSQDGYTGDVLID 193

Qy        247 LLTKYFKL-ARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIE 305
                 :||      |:  :||      :|    :  :|    :   | : |:|
Db        194 A----YKLDGTMMWRINLGK---NIRAGAHYTQFLVYDFDGDGKAEIAM---------- 235

Qy        306 QIYKTCEPTK-ALGGHAGWAPFPVR-PRKRHTSKTSYM---------HDETMDYP----- 349
                 || : ||    |   | |     |  : |  | .  |:           |::|
Db        236 ---KTADGTKDGKGKVIGNANADYRNAQGRILSGPEYLTVFKGDTGAELTTVNYEPARGN 292

Qy        350 --------------FYALTETINGSGPN---QRGKYKSAYMIK-DFPDFQIDVIWKYLTE 391
                           | |    ::|  |:    || |    ::  :|   ::  :| : ::
Db        293 VADWGDSYGNRVDRFLAGVAYLDGERPSFVMARGYYTRTVLVAYNFRGGKLTKLWTFDSD 352

Qy        392 VP-DGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEED-KDAVN 449
               | :|  : :   :|   |: |    :::: | ||   ::  |  | | |   |     |
Db        353 APGNGAYAGQGNHSLSVADVDGDGKDEIIYGAMAV---DHDGKGLYSTGWGHGDAMHTGN 409

Qy        450 LKWIR------DFYEEMYEPYG-GVPDPNT-----QVESGKGVPEGCYFNYPDVD----- 492
               |   |         :|    |||    |  |      | :|| |  |      |:|
Db        410 LDPSRPGLEVFQVHENSNSPYGLSFRDAKTGKIIWGVHAGKDVGRGM---AADIDPRYEG 466

Qy        493 LNNWKNGKYGALELYFLGN 511
                | ||     :   :||
Db        467 AEVWANGSLYTAKGVKIGN 485


RESULT 12
US-09-423-468A-13
; Sequence 13, Application US/09423468A
; Patent No. 6469149
; GENERAL INFORMATION:
;  APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnol
;  APPLICANT:  Guy Jerome Corneel Bauw
;  APPLICANT:  Mark William Davey
;  APPLICANT:  Jens Ostergaard
;  APPLICANT:  Marc Charles Ernest Van Montagu
;  TITLE OF INVENTION: PRODUCTION OF ASCORBIC ACID IN PLANTS
;  FILE REFERENCE: DECLE5.001APC
;  CURRENT APPLICATION NUMBER: US/09/423,468A
;  CURRENT FILING DATE:  2000-02-15
;  PRIOR APPLICATION NUMBER: NL 1006000
```

```
;  PRIOR FILING DATE: 1997-05-07
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 600
;   TYPE: PRT
;   ORGANISM: Brassica oleracea
US-09-423-468A-13

  Query Match             3.1%;  Score 94;  DB 4;  Length 600;
  Best Local Similarity   17.9%;  Pred. No. 1.1;
  Matches   75;  Conservative   55;  Mismatches  128;  Indels  160;  Gaps   18;

Qy        143 LGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHT 202
              :|| |  |  |  |||   |:|     :|:   ::|       : ||   || |||
Db        210 IGGIIQVGAHGTGARLP--PID----EQVIGMKLVTPAKGTIELSKD---NDPELFHLAR 260

Qy        203 GGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNT 262
              | ||  |:: :     |    |  :  : : |        |:::   : ||
Db        261 CGLGG-LGVVAEV---TLQCVERQELLEHTYVS-------TLEEIKKNHKKLL------- 302

Qy        263 VGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAG 322
                    :  : | |||  |::                    :  || |     |
Db        303 ---------STNKHVKYLYIPYTDTV-----------------VVVTCNPVSKWSG--- 332

Qy        323 WAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQI 382
                                                |  : || :   :|   |
Db        333 -------------------------------------APKDKPKYTTEEALKHVRDLYR 354

Qy        383 DVIWKYLTE-----VPDG-------LTSAEMKDALLQVDMFG----GEIHKVVWDATAVA 426
              : | ||  :      ||        |:  |::| |: :|       |::::        |
Db        355 ESIVKYRVQDSSKKTPDSREPDINELSFTELRDKLIALDPLNDVHVGKVNQ--------A 406

Qy        427 QREYIIKLQ-YQTYWQEE--DKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEG 483
              : |:  | : |:  | :|    |    :|:     |   | | :  |: :
Db        407 EAEFWKKSEGYRVGWSDEILGFDCGGQQWV----SETCFPAGTLAKPSMK---------- 452

Qy        484 CYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEI---FTNKQSIPTKP 538
                                  :| ::  |  ||: : :  |: |   :| :   |  |
Db        453 --------------------DLEYIEQLKELIQKEAIPAPSPIEQRWTGRSKSPMSP 489


RESULT 13
US-08-704-711A-17
; Sequence 17, Application US/08704711A
; Patent No. 6114159
;  GENERAL INFORMATION:
;    APPLICANT:  WILL, Horst
;    APPLICANT:  HINZMANN, Bernd
;    TITLE OF INVENTION:  DNA SEQUENCES FOR MATRIX
;    TITLE OF INVENTION:  METALLOPROTEASES, THEIR PRODUCTION AND USE
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/704,711A
;      FILING DATE:  20-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/DE95/00357
;      FILING DATE:  17-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 4438838.1
;      FILING DATE:  21-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 4409663.1
;      FILING DATE:  17-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  GRANADOS, Patricia D.
;      REGISTRATION NUMBER:  33,683
;      REFERENCE/DOCKET NUMBER:  26083/124
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  466 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-704-711A-17

  Query Match             3.1%;  Score 92;  DB 3;  Length 466;
  Best Local Similarity   20.2%;  Pred. No. 1.2;
  Matches  100;  Conservative   52;  Mismatches  158;  Indels  184;  Gaps   27;

Qy         94 INVTGLVES-----------GYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVG 142
              :||||                |  |  |: :: |:  |      :|
Db         72 LNVTGKPNEETLDMMKKPRCGVPDSGGFMLTPGNPKWERTNLTYRIRN------------ 119

Qy        143 LGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWA-- 200
                                          | |:|  | : :       |    |:
Db        120 -------------------------------YTPQLSEAEVERAI------KDAFELWSVA 143

Qy        201 ----HTGGGGGNFGIITKYYFKD----LPM-SPRGVIASNLHFSWDGFTRDALQDLLTKY 251
                  |    |  |  :| :|      |   | |::|         |   || |
Db        144 SPLIFTRISQGEADINIAFYQRDHGDNSPFDGPNGILAHAFQ-PGQGIGGDAHFD----- 197

Qy        252 FKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSND---------AEREVAQDRHYHLEA 302
                |  | ||   : :|  || ||   |  ::|:|          | || :   :| |
Db        198 ---AEETWTNTSANYNLFLVAAHEFGHSLGLAHSSDPGALMYPNYAFRETS---NYSLPQ 251

Qy        303 D----IEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETIN 358
                  |    |: ||              | :  |::|    | |     | ::   | |:|
Db        252 DDIDGIQAIY------------GLSSNPIQPTGPSTPKPC---DPSL--TFDAIT---- 289

Qy        359 GSGPNQRGK---YKSAYMIKDFPDFQ------IDVIWKYLTEVPDGLTSA-EMKDALLQV 408
                  ||:   :|  |  :  |  |      | : |  |   | |: :| |  |  | :
Db        290 ----TLRGEILFFKDRYFWRRHPQLQRVEMNFISLFWPSL---PTGIQAAYEDFDRDL-I 341

Qy        409 DMFGGEIHKVV--WD--------------ATAVAQREYIIKLQYQTYWQEEDKDAVNLKW 452
              :| |  :  :  :|               ::|   :  :  : :||:   |:      |
Db        342 FLFKGNQYWALSGYDILQGYPKDISNYGFPSSVQAIDAAVFYRSKTYFFVNDQ-----FW 396

Qy        453 IRDFYEEMYEP------YGGVPDPNTQVESGKGVFEGCYF-------NYPDVDL------ 493
                |   :  ||       |  |     :||   ||:  :|         |   ||
Db        397 RYDNQRQFMEPGYPKSISGAFPG----IESKDAVFQQEHFFHVFSGPRYYAFDLIAQRVT 452

Qy        494 -----NNWKNGKYG 502
                    | | | :||
Db        453 RVARGNKWLNCRYG 466


RESULT 14
US-09-521-220-17
; Sequence 17, Application US/09521220
; Patent No. 6399348
;  GENERAL INFORMATION:
;       APPLICANT: WILL, Horst
;                  HINZMANN, Bernd
;       TITLE OF INVENTION: DNA SEQUENCES FOR MATRIX
;                           METALLOPROTEASES, THEIR PRODUCTION AND USE
;       NUMBER OF SEQUENCES: 22
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Foley & Lardner
```

```
;           STREET: 3000 K Street, N.W., Suite 500
;           CITY: Washington
;           STATE: D.C.
;           COUNTRY: USA
;           ZIP: 20007-5109
;         COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/521,220
;           FILING DATE: 08-Mar-2000
;           CLASSIFICATION: <Unknown>
;                           21-OCT-1994
;                           17-MAR-1994
;         PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: 08/704,711
;           FILING DATE: <Unknown>
;           APPLICATION NUMBER: DE 4438838.1
;           FILING DATE: 21-OCT-1994
;           APPLICATION NUMBER: DE 4409663.1
;           FILING DATE: 17-MAR-1994
;         ATTORNEY/AGENT INFORMATION:
;           NAME: GRANADOS, Patricia D.
;           REGISTRATION NUMBER: 33,683
;           REFERENCE/DOCKET NUMBER: 26083/124
;         TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (202)672-5300
;           TELEFAX: (202)672-5399
;           TELEX: 904136
;       INFORMATION FOR SEQ ID NO: 17:
;         SEQUENCE CHARACTERISTICS:
;           LENGTH: 466 amino acids
;           TYPE: amino acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;         SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-521-220-17

  Query Match             3.1%;  Score 92;  DB 4;  Length 466;
  Best Local Similarity   20.2%;  Pred. No. 1.2;
  Matches  100;  Conservative   52;  Mismatches  158;  Indels  184;  Gaps   27;

Qy          94 INVTGLVES----------GYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVG 142
               :||||                |  |  |: :: |:  |      :|
Db          72 LNVTGKPNEETLDMMKKPRCGVPDSGGFMLTPGNPKWERTNLTYRIRN------------ 119

Qy         143 LGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWA-- 200
                                          | |:|  | : :       |    |:
Db         120 --------------------------------YTPQLSEAEVERAI------KDAFELWSVA 143

Qy         201 ----HTGGGGGNFGIITKYYFKD----LPM-SPRGVIASNLHFSWDGFTRDALQDLLTKY 251
                   |    |   |   :| :|      |   | |::|         |    ||  |
Db         144 SPLIFTRISQGEADINIAFYQRDHGDNSPFDGPNGILAHAFQ-PGQGIGGDAHFD----- 197

Qy         252 FKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSND---------AEREVAQDRHYHLEA 302
                  |   | ||   : :|  || ||   |  ::|:|         | || :   :| |
Db         198 ---AEETWTNTSANYNLFLVAAHEFGHSLGLAHSSDPGALMYPNYAFRETS---NYSLPQ 251

Qy         303 D----IEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETIN 358
               |    |: ||             | :  |::|    | |      | ::   | |:|
Db         252 DDIDGIQAIY------------GLSSNPIQPTGPSTPKPC---DPSL--TFDAIT---- 289

Qy         359 GSGPNQRGK---YKSAYMIKDFPDFQ------IDVIWKYLTEVPDGLTSA-EMKDALLQV 408
                  ||:   :|  |  :  |  |      | : |  |   | |: :| |  |  | :
Db         290 ----TLRGEILFFKDRYFWRRHPQLQRVEMNFISLFWPSL---PTGIQAAYEDFDRDL-I 341

Qy         409 DMFGGEIHKVV--WD-------------ATAVAQREYIIKLQYQTYWQEEDKDAVNLKW 452
                :| |  :  :  :|               ::|  :  :  : :||:   |:       |
Db         342 FLFKGNQYWALSGYDILQGYPKDISNYGFPSSVQAIDAAVFYRSKTYFFVNDQ-----FW 396

Qy         453 IRDFYEEMYEP------YGGVPDPNTQVESGKGVFEGCYF-------NYPDVDL------ 493
```

```
                    |   :  ||       |  |     :||   ||:  :|        |   ||
Db         397 RYDNQRQFMEPGYPKSISGAFPG----IESKDAVFQQEHFFHVFSGPRYYAFDLIAQRVT 452

Qy         494 -----NNWKNGKYG 502
                    | | | :||
Db         453 RVARGNKWLNCRYG 466


RESULT 15
US-08-652-877-12
; Sequence 12, Application US/08652877
; Patent No. 6187548
;  GENERAL INFORMATION:
;    APPLICANT:  Akerstrom, Goran
;    APPLICANT:  Juhlin, Claes
;    APPLICANT:  Rask, Lars
;    APPLICANT:  Crumley, Gregg R.
;    APPLICANT:  Morse, Clarence C.
;    APPLICANT:  Murray, Edward M.
;    APPLICANT:  Hjalm, Goran
;    TITLE OF INVENTION:  Human Calcium Sensor Protein, Fragments
;    TITLE OF INVENTION:  Thereof and DNA Encoding Same
;    NUMBER OF SEQUENCES:  106
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rhone-Poulenc Rorer Inc.
;      STREET:  500 Arcola Rd., 3C43
;      CITY:  Collegeville
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19426-0107
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  System 7.5.1
;      SOFTWARE:  Word 6.0 (Patentin)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/652,877
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/15203
;      FILING DATE:  22-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/344,836
;      FILING DATE:  23-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/487,314
;      FILING DATE:  07-JUNE-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Savitzky, Martin
;      REGISTRATION NUMBER:  29,699
;      REFERENCE/DOCKET NUMBER:  A1355E-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-454-3816
;      TELEFAX:  610-454-3808
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 944 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-652-877-12

  Query Match             3.1%;  Score 92;  DB 4;  Length 944;
  Best Local Similarity   22.3%;  Pred. No. 3.7;
  Matches   53;  Conservative   35;  Mismatches   90;  Indels   60;  Gaps   12;

Qy         325 PFPVRPRKRHTSK---TSYMHDE----TMDYPF-------YALTETINGSGPNQRGKYKS 370
               |  || || : |    : |: ||     :|| :         :  |: | |:: |  |
Db         348 PDNVRIRKYNLSSERFSEYLQDEEYIQAVDYDWDPKDIGLSVVYYTVRGEG-SRFGAIKR 406

Qy         371 AYMIKDFPDF---------QIDVIWKYLTEVPDGLTS---------AEMKDALLQVDMFG 412
               ||:    |:|         ::|:  ||: : |||:           :::|:  ::|
```

```
Db        407 AYI----PNFESGRNNLVQEVDLKLKYVMQ-PDGIAVDWVGRHIYWSDVKNKRIEVAKLD 461

Qy        413 GEIHKVVW------DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGG 466
              |   | |      | | |       : | : | | :| |  :   |:
Db        462 GRYRK--WLISTDLDQPAAIAVNPKLGLMFWTDWGKEPK--IESAWMNG---------- 506

Qy        467 VPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDP 524
                | |  |    |   |   :| : |   | : |   :|       :| : ||   :|
Db        507 -EDRNILVFEDLGWPTGLSIDYLNNDRIYWSDFKEDVIETIKYDGTDRRVIAKEAMNP 563
```

Search completed: June 10, 2003, 10:00:51
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 09:54:36 ; Search time 35 Seconds
(without alignments)
3214.335 Million cell updates/sec

Title: US-09-998-284-2
Perfect score: 2997
Sequence: 1 MATLPQKDPGYIVIDVNAGT.....IFTNKQSIPTKPLKEPKQTK 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2997	100.0	546	10	P93762	P93762 chondrus cr
2	771.5	25.7	508	16	Q8Y3D0	Q8y3d0 ralstonia s
3	736.5	24.6	685	16	Q8ZJ22	Q8zj22 yersinia pe
4	674.5	22.5	550	16	Q9EX55	Q9ex55 streptomyce
5	635.5	21.2	539	2	Q9ALN1	Q9aln1 saccharopol
6	633.5	21.1	545	16	Q93S08	Q93s08 streptomyce
7	590.5	19.7	489	2	Q8VWA5	Q8vwa5 streptomyce
8	440.5	14.7	485	2	Q9ZAR8	Q9zar8 streptomyce
9	399	13.3	480	16	P71091	P71091 bacillus su
10	368.5	12.3	530	10	Q9FZC5	Q9fzc5 arabidopsis
11	368.5	12.3	530	10	Q949N1	Q949n1 arabidopsis
12	347.5	11.6	539	10	Q9SVG3	Q9svg3 arabidopsis
13	346	11.5	528	10	Q9SVG7	Q9svg7 arabidopsis
14	346	11.5	535	10	Q9AYM8	Q9aym8 vigna ungui
15	333.5	11.1	529	10	Q9FZC8	Q9fzc8 arabidopsis
16	328.5	11.0	535	10	Q9FZC4	Q9fzc4 arabidopsis
17	324.5	10.8	541	10	Q9LPC3	Q9lpc3 arabidopsis
18	322	10.7	540	10	Q9SVG4	Q9svg4 arabidopsis
19	317.5	10.6	534	10	Q93ZA3	Q93za3 arabidopsis
20	305	10.2	542	10	Q9FKU9	Q9fku9 arabidopsis
21	304.5	10.2	552	10	Q8RWA4	Q8rwa4 arabidopsis
22	302.5	10.1	537	10	Q9FKU8	Q9fku8 arabidopsis
23	302.5	10.1	552	10	Q9FZC7	Q9fzc7 arabidopsis
24	301	10.0	542	10	Q93Y11	Q93y11 arabidopsis
25	300.5	10.0	532	10	O64743	O64743 arabidopsis
26	300	10.0	532	10	Q9SVG5	Q9svg5 arabidopsis
27	298.5	10.0	527	10	Q9FZC6	Q9fzc6 arabidopsis
28	297	9.9	533	10	Q9FI21	Q9fi21 arabidopsis
29	296	9.9	536	10	Q9SA99	Q9sa99 arabidopsis
30	294.5	9.8	541	10	Q9FKV0	Q9fkv0 arabidopsis
31	278.5	9.3	447	16	O06997	O06997 bacillus su
32	278	9.3	533	10	Q9SA89	Q9sa89 arabidopsis
33	277.5	9.3	526	10	Q9SA88	Q9sa88 arabidopsis
34	277	9.2	527	10	Q9SA85	Q9sa85 arabidopsis
35	275.5	9.2	715	10	Q9LNL9	Q9lnl9 arabidopsis
36	274	9.1	532	10	Q9FKV2	Q9fkv2 arabidopsis
37	274	9.1	536	10	Q9ZPP5	Q9zpp5 berberis st
38	273	9.1	535	10	Q9FI25	Q9fi25 arabidopsis
39	272.5	9.1	530	10	Q94KD7	Q94kd7 arabidopsis
40	272.5	9.1	530	10	Q9SUC6	Q9suc6 arabidopsis
41	272	9.1	538	10	Q8SA59	Q8sa59 helianthus
42	266	8.9	540	10	Q8SA60	Q8sa60 lactuca sat
43	252	8.4	531	10	Q9SA86	Q9sa86 arabidopsis
44	251.5	8.4	464	2	Q9KHK2	Q9khk2 streptomyce
45	249.5	8.3	540	10	O64745	O64745 arabidopsis

ALIGNMENTS

RESULT 1
P93762
ID P93762 PRELIMINARY; PRT; 546 AA.
AC P93762;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hexose oxidase.
OS Chondrus crispus (Carragheen).
OC Eukaryota; Rhodophyta; Florideophyceae; Gigartinales; Gigartinaceae;
OC Chondrus.
OX NCBI_TaxID=2769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97269074; PubMed=9111074;
RA Hansen O.C., Stougaard P.;
RT "Hexose oxidase from the red alga Chondrus crispus. Purification,
RT molecular cloning, and expression in Pichia pastoris.";
RL J. Biol. Chem. 272:11581-11587(1997).
DR EMBL; U89770; AAB49376.1; -.
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
SQ SEQUENCE 546 AA; 61899 MW; 8DDEA46D53C870B1 CRC64;

Query Match 100.0%; Score 2997; DB 10; Length 546;
Best Local Similarity 100.0%; Pred. No. 1.1e-222;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60

Qy         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
```

```
Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300

Qy        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360

Qy        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420

Qy        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480

Qy        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540

Qy        541 EPKQTK 546
              ||||||
Db        541 EPKQTK 546
```

RESULT 2
Q8Y3D0
ID   Q8Y3D0   PRELIMINARY;   PRT;   508 AA.
AC   Q8Y3D0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein RSc0050.
GN   RSC0050 OR RS01873.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646057; CAD13578.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   508 AA; 56031 MW; BD8F025998BC3C84 CRC64;

  Query Match   25.7%; Score 771.5; DB 16; Length 508;
  Best Local Similarity   35.2%; Pred. No. 4.2e-51;
  Matches 196; Conservative   74; Mismatches 202; Indels   85; Gaps   17;

```
Qy         12 IVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDF---VYVVYTPQGACTALDRAMEKCS 68
              : :||      : |||  ::: ||| ||  ::      : :          || | ::
Db          1 MTLDVT-----RQDPRYDTLRHGFNLRWPSSDAQAAGRIALCEKTDEVAPALQRIIDTGM 55

Qy         69 PGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWGSFKTLFRD 128
               |||  |||||||||| :     |:::: |       |    | :|  ||     |::
Db         56 RPTVR--SGGHCYEDFVSNNPGGVIVDLSLLNAPEVRADGAVRVPAGMQNWNGALELYKR 113

Qy        129 HGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHK 188
              ||  |||||||||| |||| ||| |:|:|| || ||||| |::|
Db        114 HGVTLPGGSCYSVGAGGHISGGGYGLLSRLQGLTVDWLSAVDIVT-------------V 159

Qy        189 DSEG----------NDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDG 238
              | :|           | :|| |  | |||||||| | |  || :|| |  : : | |
Db        160 DKQGRAAPRTVDAARDPDLFRACRGAGGGNFGIITSYTFARLPEAPREVALATVAFDWAA 219

Qy        239 FTRDALQDLLTKY--FKLARCDWKNTVGKFQIF---HQAAEEFVMYLYTSYSNDAEREVA 293
              | :   :||  |  :   |    :| | | :    |::| : ||   | : |
Db        220 MTPERFAELLRLYGDYWATRGKDPDTWGLFSLLKLTHKSAGQIVM--LTQFCNPD----G 273

Qy        294 QDRHYHLEADIEQIYKTCEPTKALGGHAGWAPF------PVRPRKRHTSKTSYMHDETMD 347
              |   :  |   :: | |  |      |: |       :   | ||         :
Db        274 TCRDLSVLNDFLARFQVCAPVPARARPPGYGPAHRQGVGQLLCSKPHT---------VVQ 324

Qy        348 YPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQ 407
              | :  |:::|||||||||||||| : |   :  |: :||    |:   ::  :|||
Db        325 YDWLTATQSLNGSGPNQRGKYKSAYMKRGFTAREAQRIYTHLTRTVQGI---DLSQSLLQ 381

Qy        408 VDMFGGEIHKVVWDA-TAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMY----- 461
              || :|| ::|    | ||| ||  ::|||||||      || :|:|| ||| ::|
Db        382 VDSYGGAVNKAERIADTAVPQRASVMKLQYQTYWTSAADDAGHLRWIGDFYRDVYGTPDV 441

Qy        462 -EPYGGVPDPNTQVESGKGVFEGCYFNYPDVDL---NNWKNGKYGALELYFLGNLNRLIK 517
               |: | | |  :       :|||| ||||||:     |    ||  :||       | :
Db        442 SAPHAGTPYPGDR-------YEGCYINYPDVDMLAYPFWPQLYYGDGDLYAF-----LQR 489

Qy        518 AKWLWDPNEIFTNKQSI 534
              |  :||| || :  |:
Db        490 VKRRYDPNNIFHHAMSV 506

RESULT 3
Q8ZJ22
ID   Q8ZJ22      PRELIMINARY;      PRT;   685 AA.
AC   Q8ZJ22;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative oxidoreductase.
GN   YPO0308.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414142; CAC89170.1; -.
DR   InterPro; IPR003610; CBM_5_12.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF02839; CBM_5_12; 1.
DR   Pfam; PF01565; FAD_binding_4; 1.
DR   Pfam; PF00041; fn3; 1.
DR   SMART; SM00060; FN3; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   685 AA;  75266 MW;  597853869F657CC6 CRC64;

  Query Match             24.6%;  Score 736.5;  DB 16;  Length 685;
  Best Local Similarity   33.6%;  Pred. No. 3.2e-48;
  Matches  195;  Conservative   77;  Mismatches  205;  Indels  103;  Gaps   20;
```

```
Qy         13 VIDVNAGTPDKPDPRLPSMKQGFNRRW---IGTNIDFVYVVYTPQGACTALDRAMEKCSP 69
Db          3 IIDKNVSTYE-------TLQKGFNLRWPPNVEQGAETIYICTTPDEVFAATNTALAAGNR 55

Qy         70 GTVRIVSGGHCYEDFVFDECVKAIINVTGLVE-SGYDDDRGYFVSS-------------- 114
Db         56 ITVR--SGGHCYEGFVSNKLSTERLSIIDLGEMSGLDYDEDKTITSLWDANKNTYRFKSL 113

Qy        115 -GDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVV 173
Db        114 TGNQNWNGYVSLYKRSGRTIPGGSCYSVGVGGHISGGGYGLLSRLHGLTVDWVTGVDILV 173

Qy        174 KPVLTEDSVLKYVHKDSEGN-DGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNL 232
Db        174 PVGNAHRLAFRHVRADSVSEVDRELLMACCGAGGGNFGIIIAYYFDDLPKAPQKAYWIPL 233

Qy        233 HFSWDGFTRDALQDLLTKYFKLARCDW-----------KNTVGKFQIF------H-QAAE 274
Db        234 TYPW-----SSLKATFPAFLK-AYWQWFADNDVNATSTKEGVGNGGLFTLLKLNHIDASD 287

Qy        275 EFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAP-------FP 327
Db        288 NVVLAIQYTGPNG---QVGGANDIPLNDFIEK----------MNAAAGMTPTIYDDFILP 334

Qy        328 VRPRKRHT---SKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDV 384
Db        335 NIPPFKHLYPGRKIGRTVDESASMDWLHVTQMINGSGSNQRGKYKSDYQIKQFSD---EM 391

Qy        385 IWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEED 444
Db        392 CHALLTHLTTATADKRFNQSLVQIDSYGGAINSRGIGATAVSQRNSLLKAQYQTYWTNEA 451

Qy        445 KDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLN---------N 495
Db        452 DDQTHLTWIRNIYAAV---HNGKPAPPE--------FEGCYINYPDIDMKYTDSGBEDPN 500

Qy        496 WKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIP 535
Db        501 WLNLYYG----WDTQLIKRLIALKARIDPNNIFHHELSIP 536


RESULT 4
Q9EX55
ID   Q9EX55      PRELIMINARY;      PRT;   550 AA.
AC   Q9EX55;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative secreted FAD-binding protein.
GN   SCO6272 OR SC2C4.02.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL512902; CAC22143.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
SQ   SEQUENCE   550 AA;  59895 MW;  EB4A23E0F58E4ABC CRC64;

  Query Match             22.5%;  Score 674.5;  DB 16;  Length 550;
  Best Local Similarity   33.2%;  Pred. No. 1.4e-43;
  Matches  185;  Conservative   82;  Mismatches  180;  Indels  111;  Gaps   19;

Qy         23 KPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTALDRAMEKCSPGTVRIVSGGHCYE 82
Db         62 RSDPRYQGLISGSNQRWVG-NPDYIRVVSSAGQVVRAVQEAVDKGLKIAVR--SGGHCDE 118

Qy         83 DFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWGS-FKTLFRDHGRVLPGGSCYSV 141
Db        119 DFVANRDVRVVIDMAGMDSVTYDRERRAFAVGPGARLGTVYRTLYKRWGVVLPGGTCPTV 178

Qy        142 GLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGN-------- 193
Db        179 GAGGHITGGGYGALSRSRGLTVDHLYAVEV--------------VHVDARGRARKVVATR 224

Qy        194 -----DGELFWAHTGGGGGNFGIITKYYFKD-----------LPMSPRGVIASNLHFSWD 237
Db        225 EEDDPNRELWWAHTGAGGGNFGVITRYWLRSPDAVGDDPSTLLPAAPSEVLLSDVSWSWD 284

Qy        238 GFTRDALQDLLTKYFKL-ARCDWKNTVGKF-----QIFHQAAEEFVMYLYTSYSNDAERE 291
Db        285 DLDEASFTRLLRNFTEWHARNSAPDSPGRFLFSQLKTMHKAAGYFRM----STQVDA--- 337

Qy        292 VAQDRHYHLEADIEQIYKTCEPTKALGG--HAGWAPFPVRPRKRHTSKTSYMHDETMDYP 349
Db        338 AAPDADRLLDDYLAAISEGTGVTYHVGDRYRAPW------------------------- 371

Qy        350 FYALTE---TINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTE----VPDGLTSAEMK 402
Db        372 LYAVTEWSGFVEASVP--RWKSKSAYVREVMPEEQLRAVYRQLTRDDYPGPYG------- 422

Qy        403 DALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYE 462
Db        423 --MIAIVGFGGKINEVAPGDTATAQRDSIAKMLYCSLWSDPADDALHQRWIREAYEDVYA 480

Qy        463 PYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGA--LELYFLGNLNRLIKAKW 520
Db        481 STGGVPRPG-------GVNDGCYINYADADLADPALNRSGIPWHELYFKGNYPRLQRVKA 533

Qy        521 LWDPNEIFTNKQSIPTKP 538
Db        534 TWDPRNVFSHRLGI-RKP 550


RESULT 5
Q9ALN1
ID   Q9ALN1      PRELIMINARY;      PRT;   539 AA.
AC   Q9ALN1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

```
DE   Putative oxidoreductase.
GN   SPNJ.
OS   Saccharopolyspora spinosa.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=60894;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257765; PubMed=11358695;
RA   Waldron C., Matsushima P., Rosteck P.R. Jr., Broughton M.C.,
RA   Turner J., Madduri K., Crawford K.P., Merlo D.J., Baltz R.H.;
RT   "Cloning and analysis of the spinosad biosynthetic cluster of
RT   Saccharopolyspora spinosa.";
RL   Chem. Biol. 8:487-499(2001).
DR   EMBL; AY007564; AAG23271.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
DR   PROSITE; PS00862; OX2_COVAL_FAD; UNKNOWN_1.
SQ   SEQUENCE   539 AA;  58473 MW;  C4AF114B40477CAE CRC64;

  Query Match             21.2%;  Score 635.5;  DB 2;  Length 539;
  Best Local Similarity   30.2%;  Pred. No. 1.4e-40;
  Matches  174;  Conservative   91;  Mismatches  216;  Indels   95;  Gaps   16;

Qy          2 ATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTALD 61
              | : :: ||  :      | :  ||| | :  | | |: |   : :::  : :    |:
Db         25 APMNRRTPGTEI------TVEPDDPRYPDLVVGHNPRFTG-KPERIHIASSAEDVVHAVA 77

Qy         62 RAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYF-VSSGDTNWG 120
               |:   :   | : |||||:|: | |  :: ::::: |    ||  || | : :|
Db         78 DAVR--TGRRVGVRSGGHCFENLVADPAIRVLVDLSELNRVYYDSTRGAFAIEAGAALGQ 135

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ::|||:: |  :| |:|  || ||||:||| | |:|  |  ||:| ||||||     |
Db        136 VYRTLFKNWGVTIPTGACPGVGAGGHILGGGYGPLSRRFGSVVDYLQGVEVVVVDQAGEV 195

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKD----------LPMSPRGVIA 229
              :::   ::| |   :|:|||||||||||||:|::: :            ||  |  |:
Db        196 HIVE-ADRNSTGAGHDLWWAHTGGGGGNFGIVTRFWLRTPDVVSTDAAELLPRPPATVLL 254

Qy        230 SNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSY----- 284
              : |: |   |  :   ||  :              | :   | | |   : |:::
Db        255 RSFHWPWHELTEQSFAVLLQNF-----------GNWYEQHSAPESTQLGLFSTLVCAHR 302

Qy        285 --------------SNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRP 330
                           :||| :|:    || |   |:  |  | : |     |
Db        303 QAGYVTLNVHLDGTDPNAERTLAE----HLSAINAQVGVT--PAEGLRETLPWL------ 350

Qy        331 RKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLT 390
              | |            |: |   |    || | |:||:     : |:  ::: ||
Db        351 --RSTQVAG-----------AIAE--GGEPGMQRTKVKAAYLRTGLSEAQLATVYRRLT 394

Qy        391 EVPDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNL 450
                  :   :   | | :  :||  : |   |||:|||: ::|  : | | |  :|  :|
Db        395 -----VYGYDNPAAALLLLGYGGMANAVAPSATALAQRDSVLKALFVTNWSEPAEDERHL 449

Qy        451 KWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNN--WKNGKYGALELYF 508
              |||  || |||   ||||  | |:|       :| | ||||  || :  |         :||:
Db        450 TWIRGFYREMYAETGGVPVPGTRV-------DGSYINYPDTDLADPLWNTSGVAWHDLYY 502

Qy        509 LGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLKEPKQ 544
              |  || :||  |||  || :  ||       | |
Db        503 KDNYPRLQRAKARWDPQNIFQHGLSIKPPARLSPGQ 538


RESULT 6
Q93S08
ID   Q93S08      PRELIMINARY;      PRT;   545 AA.
AC   Q93S08;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative FAD-binding protein.
GN   SCO6281 OR SC1G7.07.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K., Harris D.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL591083; CAC37882.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
SQ   SEQUENCE   545 AA;  58522 MW;  4D72C37A5433B76A CRC64;

  Query Match             21.1%;  Score 633.5;  DB 16;  Length 545;
  Best Local Similarity   32.9%;  Pred. No. 2e-40;
  Matches  176;  Conservative   81;  Mismatches  205;  Indels   73;  Gaps   18;

Qy         25 DPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTALDRAMEKCSPGTVRIVSGGHCYEDF 84
              |||  ::  | | || ||  |:| |          |:  : |     |:  |||||||||
Db         56 DPRYATLVAGANGRWKGTP-DYVLVASDADQVVQAVRETLAKGLRFAVK--SGGHCYEDF 112

Qy         85 VFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWGS-FKTLFRDHGRVLPGGSCYSVGL 143
               :  |: :|::: :    :|  |  |         || :: |:   |  ||||:| || :
Db        113 TTNSGVRVLIDISAMAAIEFDASRRAFAIGPGAQLGSVYQKLYDGWGVTLPGGTCPSVAV 172

Qy        144 GGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYV-HKDSEGNDGELFWAHT 202
              |||| ||| | ||| ||: ||:|  :|:||  |    :| | |  :: :  : ||:|||
Db        173 GGHIPGGGYGPLARSHGITVDYLYAIEIVV--VDRAGTVRKIVATREHDDPNRELWWAHT 230

Qy        203 GGGGGNFGIITKYYFKD----------LPMSPRGVIASNLHFSWDGFTRDALQDLLTKYF 252
              | ||||:||:|:|:|::          || :|| :| | : | ||    :|    | : |
Db        231 GAGGGNYGIVTRYWFRNDAGGLDPAALLPRAPRELIISEVTFPWDNTMTEAAFSRLLRNF 290

Qy        253 KL-----ARCD--WKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIE 305
                    |  |  :       :  |::| ||:|     |  ||     |      |:| :
Db        291 SAWHVANASADHPYARLFSALKPRHRSAGEFLM----STQIDAAVPGADAL---LDAYLA 343

Qy        306 QIYKTCEPTKALGGHAGWA-PFPVRPRKRHTSKTSYMHDETMDYPFYALT-ETINGSGPN 363
              :|             ||    : |  |||            :|: :  |    : | |
Db        344 EIV-----------AGTGLTYTVVTRKR-----------VDWLYNVLNWPGLGGDGFE 379
```

```
Qy        364 QRGKY--KSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVWD 421
              :|::  ||||: |  || ||   :|:||      |  :   ||::::  :|   :
Db        380 GKGRFKAKSAYLRKTLPDAQIKAFYKHLTR-----TDYDNPAALVEIAGYGAAANLPASW 434

Qy        422 ATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGVF 481
              ||| |||: :||: :   |  | :|  || |:|:|| :::   |||| |:        ||
Db        435 ATATAQRDSVIKMLFVNLWATEAEDRRNLAWVREFYRDVFAATGGVPRPS-------GVN 487

Qy        482 EGCYFNYPDVDLNN--WKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSI 534
              :| : || | || :        |   |||     ||  ||  |||  :||:   |
Db        488 DGAFINYADADLADPALNTSGIGWNTLYFKDGYCRLQAAKTQWDPRNVFTHALGI 542


RESULT 7
Q8VWA5
ID   Q8VWA5      PRELIMINARY;      PRT;   489 AA.
AC   Q8VWA5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AclO.
GN   ACLO.
OS   Streptomyces galilaeus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33899;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3AR-33;
RA   Chung J., Fujii I., Tsukamoto N., Sankawa U., Ebizuka Y.;
RT   "Aklavinone-aclacinomycin biosynthesis gene cluster from Streptomyces
RT   galilaeus.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB008466; BAB72054.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
SQ   SEQUENCE   489 AA;  53652 MW;  BA006ADAA0BA33E7 CRC64;

  Query Match             19.7%;  Score 590.5;  DB 2;  Length 489;
  Best Local Similarity   29.3%;  Pred. No. 3.6e-37;
  Matches  159;  Conservative   91;  Mismatches  204;  Indels   89;  Gaps   17;

Qy         22 DKPDPRLPSM-KQGFNRRWIGTNIDFVYVVYTPQGACTALDRAMEKCSPGTVRIVSGGHC 80
              |: | |    :  :||| |: |    | ||||:|      |:::|:   :   : : |||||
Db          2 DRADRRYQDLVTRGFNGRFRG-RPDVVYVVHTADQVVDAVNQAV--AAGRRIAVRSGGHC 58

Qy         81 YEDFVFDECVKAIINVTGLVESGYDD-DRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCY 139
              :| ||  |  |:|:|::: : :  ||    | : |  | |   ::: |: | |  :| | |
Db         59 FEGFVDDPAVRAVIDMSQMRQVFYDSGKRAFAVEPGATLGETYRALYLDWGVTIPAGVCP 118

Qy        140 SVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGND--GEL 197
              ||:|||::|||  |  |:|   |:  | |  ||||    |:      : |   |  :|   ||
Db        119 QVGVGGHVLGGGYGPLSRRDGVVADHLYAVEVV---VVDASGQARKVVATSAADDPNREL 175

Qy        198 FWAHTGGGGGNFGIITKYYFKD-----------LPMSPRGVIASNLHFSWDGFTRDALQD 246
              :|||||||||||||:|:|:|:           || :|   :   : : |   | :|
Db        176 WWAHTGGGGGNFGIVTRYWFRTPGATGTDPSALLPKAPTSTLRHIVTWDWSALTEEAFTR 235

Qy        247 LLTKY------FKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300
              ::  :      |   : :    | :  :|| : :: : |    |    :  |
Db        236 IIDNHGAWHQRNSAAGTPYASMHSVFYLNSRAAGQILLDIQTDGGLDGAETLLND----F 291

Qy        301 EADIEQ-------IYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYAL 353
              | : :       : :| ||          |          | |    :
Db        292 VAAVNEGTGVEPAVQRTTEP---------WL--------RATLANKF------------- 321

Qy        354 TETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGG 413
                 :|  | | | ||: | :    |   ::::|:   |      |     : :  :||
Db        322 -----DTGGFDRTKSKGAYLRKPWTAAQAATLYRHLS--ADSQVWGE-----VSLYSYGG 369

Qy        414 EIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQ 473
              ::: |   ||| |||: |||:     | :   |  || |||: | |::   |||| |: :
Db        370 KVNSVPETATATAQRDSIIKVWMSATWMDPAHDDANLAWIREIYREIFATTGGVPVPDDR 429

Qy        474 VESGKGVFEGCYFNYPDVDLNN--WKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNK 531
                    || : ||||||| :  |          ||: ||  || : |  ||| ::| :
Db        430 T-------EGTFINYPDVDLADPRWNTSGVPWYTLYYKGNYPRLQRVKARWDPRDVFRHA 482

Qy        532 QSI 534
               |:
Db        483 LSV 485


RESULT 8
Q9ZAR8
ID   Q9ZAR8      PRELIMINARY;      PRT;   485 AA.
AC   Q9ZAR8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DnrW.
GN   DNRW.
OS   Streptomyces peucetius subsp. caesius.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=55158;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 27952;
RA   Hong Y.-S., Hwang C.K., Lee J.J.;
RT   "Regulation of anthracycline production in Streptomyces peucetius by
RT   the dnrW.";
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80222; AAD00354.1; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ   SEQUENCE   485 AA;  52227 MW;  07DC310677034557 CRC64;

  Query Match             14.7%;  Score 440.5;  DB 2;  Length 485;
  Best Local Similarity   28.3%;  Pred. No. 1.3e-25;
  Matches  135;  Conservative   80;  Mismatches  183;  Indels   79;  Gaps   15;

Qy         25 DPRLPSM-KQGFNRRWIGTNIDFVYVVYTPQGACTALDRAMEKCSPGTVRIVSGGHCYED 83
              ||| | :  :| | |:   : | | || | : |  |:  ::   :   ||  :||||:|
Db         49 DPRYPDLVGRGINARFT-PDPDTVRVVATAEQAVRAVQDSVRDGTRLAVR--TGGHCFES 105

Qy         84 FVFDECVKAIINVTGLVESGYDDDRGYF-VSSGDTNWGSFKTLFRDHGRVLPGGSCYSVG 142
              | |  | :|:|: :   :| :  | | || |    :::|:       :| | |  ||
Db        106 LVDDPAVTTVIDVSEMRSVYFDGELNAFSVDSGATLGTMYRSLYLGWDVTVPAGRCPEVG 165

Qy        143 LGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGND--GELFWA 200
              :|||| ||| | |:| :||  || | |||||    |:      : |    | :|   :|:||
Db        166 VGGHIAGGGGGALSRTYGLSVDHLHGVEVV---VVDSGGRARLVRATREPSDPHRDLWWA 222

Qy        201 HTGGGGGNFGIITKYYFKD---------------LPMSPRGVIASNLHFSWDGFTRDALQ 245
              ||||| |:||::|:| |:                ||  |  |:   : : ||      |
Db        223 HTGGGAGSFGLVTRYLFRSPGTDAGSRPADPGRLLPRPPGSVLRKTVRWDWDSVDETAFL 282

Qy        246 DLLTKYFKLARCDWKNTVGKFQIFHQ---------AAEEFVMYLYTSYSNDAEREVAQDR 296
              |                 |  | :|:         |  :  : |  :        | | |
Db        283 TL---------------VRNFGTWHERHAAPGDPGARLDNSLALPRTGGGPLTLETAVDA 327

Qy        297 HYHLEADIEQIYKTC--EPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALT 354
                 | |::     | ::::|          ||    |:           |: | |
Db        328 ---TRPDAEELTDRFIREVSRSVG---------ARPEVSVTT-----------LPWLAAT 364

Qy        355 ETING-SGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGG 413
                 :  :|    | | |:|::   : : |  :::: ||:        |     | : :   ||
Db        365 LVPDEFAGIKGRFKSKAAFLRTGWSERQARMVYQRLTDTSGYHNPA----ATVYLLSHGG 420

Qy        414 EIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDP 470
              |:::     ||:| |: ::|  :  :| :| :||::| |:|  | | :   ||||||
Db        421 EVNRPGPADTAMAHRDAVLKTYWSVFWFDEREDALHLDWVRASYAEFFGDAGGVPDP 477
```

```
RESULT 9
P71091
ID   P71091      PRELIMINARY;      PRT;   480 AA.
AC   P71091;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical 54.4 kDa protein (YGAK protein).
GN   YGAK.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Cummings N.J., Connerton I.F.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 114-480 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE OF 114-480 FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z82044; CAB04812.1; -.
DR   EMBL; Z99108; CAB12708.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
KW   Complete proteome.
SQ   SEQUENCE   480 AA;  54477 MW;  B43CD118F7FE4C8B CRC64;

  Query Match             13.3%;  Score 399;  DB 16;  Length 480;
  Best Local Similarity   26.0%;  Pred. No. 2e-22;
  Matches  144;  Conservative   66;  Mismatches  192;  Indels  152;  Gaps   21;

Qy         20 TPDKPDPRLPSMKQGFNRRWIGTNI------DFVYVVYTPQGACTALDRAMEKCSPGTVR 73
              | | ||          :|      |:      | :      | |  ||  | |   |   |
Db         41 TRDDPD---------YNEARTNINLSLERYPDIIVFCQNKQDALNALKWARENRVP--FR 89

Qy         74 IVSGGHCYEDF-------VFD--ECVKAIINVTGLVESGYDDDRGYFVSSGDTNWGSFKT 124
              | | ||:|         | |  |  |  :|          | :  :: :|      ::|
Db         90 IRGGRHSYENFSLLNNGLVIDLSEMKKITVN---------QDKKLAYIEAGAELGEVYRT 140

Qy        125 LFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLK 184
              |:: :|  || |:  :||| |  :||| |:| |  ||  | |  :|::|
Db        141 LWQ-YGLTLPAGTIANVGLTGLTLGGGIGLLTRAAGLTCDSLVQLEMIVA---------- 189

Qy        185 YVHKDSEG---------NDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFS 235
                : ||          |  :|||| |||||||||:|   || :|:|   :   :: :
Db        190 ---DEKEGADLITVSCSNHPDLFWASQGGGGGNFGIVTSMTFKAVPISQVSIF--SITWG 244

Qy        236 WDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQD 295
              || |                |:|                :  |::|
Db        245 WDDFEE------------VYNTWQN-----------------WPPYTDD-------- 264

Qy        296 RHYHLEADIEQIYKTCEPTKALGGHAGWAPFP-------VRPRKRHTSKTSYMHDETMDY 348
                | : ||   |     :|||   |    |       ::|  :  | ||| |   |
Db        265 ---RLTSSIEFWPKEVNRIEALGQFVG----PKTELKKLLKPLLKAGSPTSGMVKTT--- 314

Qy        349 PFYALTETINGSGPNQRGKYK--SAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALL 406
              ||       |  | ||  | |   ::: |   :  |  | :|   |:       ::| :
Db        315 PFIEAVTFFNSPGGNQPQKMKRSGSFIEKPLSERAISTIKHFLEHAPN-------QNASV 367

Qy        407 QVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGG 466
                ||   :|  | ||   |: ||  :| | |    :   |::||      : :
Db        368 WQQALGGAAGRVAPDQTAFYYRDAIIAQEYLTNWTSPGEKRQNVRWIEGLRTSLSKE--- 424

Qy        467 VPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNE 526
                          | | |:||::: ||    ||      |: || : |  :||
Db        425 -------------TMGDYVNWPDIEIRNWPRTYYGE------NVERLRRVKTTYDPEN 463

Qy        527 IFTNKQSIPTKPLK 540
              :|  :|||| ||:
Db        464 VFRFEQSIP--PLR 475


RESULT 10
Q9FZC5
ID   Q9FZC5      PRELIMINARY;      PRT;   530 AA.
AC   Q9FZC5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   T1K7.23 protein.
GN   T1K7.23.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Sakano H., Liu S.X., Yu G., Lee J.M., Lenz C., Pham P., Toriumi M.,
RA   Chin C., Chiou J., Choi E., Chung M., Gonzalez A., Howng B., Liu A.,
RA   Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
RA   Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA   Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
RA   Ecker J.R., Federspiel N.A., Theologis A.;
RT   "The sequence of BAC T1K7 from Arabidopsis thaliana chromosome 1.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC013427; AAF98577.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
SQ   SEQUENCE   530 AA;  58947 MW;  116C6CA06BE4C322 CRC64;

  Query Match             12.3%;  Score 368.5;  DB 10;  Length 530;
  Best Local Similarity   27.0%;  Pred. No. 5.3e-20;
  Matches  134;  Conservative   76;  Mismatches  180;  Indels  107;  Gaps   23;

Qy         72 VRIVSGGHCYEDFVFDECVKAII-------NVTGLVESGYDDDRGYFVSSGDTNWGSFKT 124
```

```
Db        103 LRIRSGGHDYEGLSYMSSVPFVILDMYNLRSITVDVSS-----KKAWIQAGAT-LGELYT 156

Qy        125 LFRDHGRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSV 182

Db        157 NVNDVSQTLAFPAGVCATVGAGGHISGGGYGNLMRKYGITVDHVIDAQII-------DVN 209

Qy        183 LKYVHKDSEGNDGELFWAHTGGGGGNFGIIT--KYYFKDLPMSPRGVIASNLHFSWDGFT 240

Db        210 GKLLNRATMGED--LFWAIRGGGGGSFGVILSWKINLVDVPKIVT-VFKVNKTLEQGG-- 264

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEE----------FVMYLYTSYSND 287

Db        265 ----TDVLYK--------WQLVASKFPESLFVRAMPQVANGTKRGERTITVVFYAQFLGR 312

Qy        288 AEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMD 347

Db        313 TDALMAIMNQNWPELGLK--HEDCQEMSWLNSTLFWADYPA------GTPTSIL----LD 360

Qy        348 YPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQ 407

Db        361 RP----------SSPGDFFKSKSDYVKKPIPKEGLEKLWKTMLKFNNNI-------VWMQ 403

Qy        408 VDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGG 466

Db        404 FNPYGGVMDRIPATATAFPHRKGNLFKIQYFTTWFNANATMSSLSQMKELY-EVAEPY-- 460

Qy        467 VPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGK--------YGALELYFLGNLNRLIKA 518

Db        461 -------VSSNP---REAFFNYRDIDVGSNPSGETNVDEAKIYGS--KYFLGNLKRLMDV 508

Qy        519 KWLWDPNEIFTNKQSIP 535

Db        509 KAKYDPDNFFKNEQSIP 525
```

RESULT 11
Q949N1
ID Q949N1 PRELIMINARY; PRT; 530 AA.
AC Q949N1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 58.9 kDa protein.
GN T1K7.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene T1K7.23 (GI:9797759).";
RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY051000; AAK93677.1; -.
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
KW Hypothetical protein.
SQ SEQUENCE 530 AA; 58947 MW; 1BC6C456BDE9A3F4 CRC64;

Query Match 12.3%; Score 368.5; DB 10; Length 530;
Best Local Similarity 27.0%; Pred. No. 5.3e-20;
Matches 134; Conservative 76; Mismatches 180; Indels 107; Gaps 23;

```
Qy         72 VRIVSGGHCYEDFVFDECVKAII-------NVTGLVESGYDDDRGYFVSSGDTNWGSFKT 124
```

```
Db        103 LRIRSGGHDYEGLSYMSSVPFVILDMYNLRSITVDVSS-----KKAWIQAGAT-LGELYT 156

Qy        125 LFRDHGRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSV 182

Db        157 NVNDVSQTLAFPAGVCATVGAGGHISGGGYGNLMRKYGITVDHVIDAQII-------DVN 209

Qy        183 LKYVHKDSEGNDGELFWAHTGGGGGNFGIIT--KYYFKDLPMSPRGVIASNLHFSWDGFT 240

Db        210 GKLLNRATMGED--LFWAIRGGGGGSFGVILSWKINLVDVPKIVT-VFKVNKTLEQGG-- 264

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEE----------FVMYLYTSYSND 287

Db        265 ----TDVLYK--------WQLVASKFPESLFVRAMPQVANGTKRGERTITVVFYAQFLGR 312

Qy        288 AEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMD 347

Db        313 TDALMAIMNQNWPELGLK--HEDCQEMSWLNSTLFWADYPA------GTPTSIL----LD 360

Qy        348 YPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQ 407

Db        361 RP----------SSPGDFFKSKSDYVKKPIPKEGLEKLWKTMLKFNNNI-------VWMQ 403

Qy        408 VDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGG 466

Db        404 FNPYGGVMDRIPATATAFPHRKGNLFKIQYFTTWFNANATMSSLSQMKELY-EVAEPY-- 460

Qy        467 VPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGK--------YGALELYFLGNLNRLIKA 518

Db        461 -------VSSNP---REAFFNYRDIDVGSNPSGETNVDEAKIYGS--KYFLGNLKRLMDV 508

Qy        519 KWLWDPNEIFTNKQSIP 535

Db        509 KAKYDPDNFFKNEQSIP 525
```

RESULT 12
Q9SVG3
ID Q9SVG3 PRELIMINARY; PRT; 539 AA.
AC Q9SVG3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Reticuline oxidase-like protein.
GN F21C20.190 OR AT4G20840.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Pohl T., Weizenegger T., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL080254; CAB45850.1; -.
DR EMBL; AL161553; CAB79084.1; -.
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
SQ SEQUENCE 539 AA; 60143 MW; 6645E9376E216426 CRC64;

```
  Query Match             11.6%;  Score 347.5;  DB 10;  Length 539;
  Best Local Similarity   26.2%;  Pred. No. 2.3e-18;
  Matches  141;  Conservative   76;  Mismatches  181;  Indels  141;  Gaps   28;

Qy         52 TPQGACTALDR------AMEKCSPG---TVRIVSGGHCYE--DFVFDECVKAIINVTGLV 100
              ||: |      |      |   ||      ::| |||| ||   :: |:     |:::: |
Db         83 TPKPAIIVTPRSDIHVSAAVTCSKSLNFLLKIRSGGHDYEGLSYISDKPF-FILDMSNLR 141

Qy        101 ESGYD-DDRGYFVSSGDT-------NWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGD 152
              :   |  |:  ::|:| |       |   |     ||   | | | :||:|||| |||
Db        142 DVSVDIADQSAWISAGATLGEVYYRIWEKSKV----HG--FPAGVCPTVGVGGHISGGGY 195

Qy        153 GILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGII 212
              | : |  || || |   ::|       |   : : : | | |  |||| :|||| :||::
Db        196 GNMLRKFGLSVDNLIDAKIV-------DVNGQILDRKSMGED--LFWAISGGGGASFGVV 246

Qy        213 TKYYFKDLPMSPRGVI----------ASNLHFSWDG---------FTRDALQDLLTKYFK 253
              |  | :|:     :          | ::   |             | |  :| :  |  |
Db        247 LGYKVKLVPVPETVTVFRVEKYMDSGAVDMVHKWQSVGPKTDRNLFLRMLIQPVTRKKVK 306

Qy        254 LARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEP 313
              |     ||    :|  ||| |  |  :                 |  :::  : |
Db        307 TVRA----TV--VALFLGRAEEVVALLGKEFP-------------ELSLKK--ENCSE 343

Qy        314 TKALGGHAGW----APFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYK 369
                  |      |  : |:          |  :|   :              || |
Db        344 MTWFQSALWWDNRVNPTQIDPK--------VFLDRNLDRANF-------------GKRK 381

Qy        370 SAYMIKDFPDFQIDVIWKYLTEVPD-GLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQR 428
              | |:  : |   |: ::| :||:   ||          : :||:: :|  :||     |
Db        382 SDYVASEIPRDGIESLFKKMTELGKIGLV----------FNPYGGKMAEVTVNATPFPHR 431

Qy        429 EYIIKLQYQTYWQEEDKDAVNLKWIRDFYEE---MYEPYGGVPDPNTQVESGKGVFEGCY 485
              : |:||   |||   ::| ::  : |  :   :|    |    | :           |
Db        432 SKLFKIQYSVTWQE---NSVEIE--KGFLNQANVLYSFMTGFVSKNPR---------NAY 477

Qy        486 FNYPDVDL-------NNWKNGK-YGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPT 536
                || |||:,      |::: |: ||    ||  | :||:| |   ||:  | |:|||||
Db        478 LNYRDVDIGVNDHGTNSYEEGEVYG--RKYFGDNFDRLVKVKTAADPDNFFRNEQSIPT 534


RESULT 13
Q9SVG7
ID   Q9SVG7      PRELIMINARY;      PRT;   528 AA.
AC   Q9SVG7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Hypothetical 58.8 kDa protein.
GN   F21C20.150 OR AT4G20800.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
RA   Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Pohl T., Weizenegger T., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL080254; CAB45846.1; -.
DR   EMBL; AL161553; CAB79080.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   528 AA;  58846 MW;  D70E7667AAAE3E7D CRC64;

  Query Match             11.5%;  Score 346;  DB 10;  Length 528;
  Best Local Similarity   27.0%;  Pred. No. 2.9e-18;
  Matches  134;  Conservative   68;  Mismatches  187;  Indels  108;  Gaps   23;

Qy         72 VRIVSGGHCYEDFVFDECVK-AIINVTGLVESGYDDDRGY-FVSSGDTNWGSFKTLFRDH 129
              :|| |||| ||   |   |   |:::  |     |  |  :| :| |  |   |
Db        102 IRIRSGGHDYEGLSFTSSVPFVILDMHDLRSITIDVFRKQAWVDAGAT-MGELYTKIAAA 160

Qy        130 GRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVH 187
              : |   || | ::| |||| ||| | | | :|: || :    :|
Db        161 SKTLAFAGGVCPTLGAGGHISGGGYGNLIRKYGISVDHVVDARIV--------------- 205

Qy        188 KDSEGN-------DGELFWAHTGGGGGNFGIIT--KYYFKDLPMS----------PRGVI 228
              |  ||          :| ||  |||| :||:|   |    |:| :           :||
Db        206 -DVNGNILTGATLGRDLLWAIRGGGGASFGVILSWKINLVDVPKTVTVFKVNKTLEQGV- 263

Qy        229 ASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDA 288
              ::: : |   :    |||  :          |   |         : :   :  | :   |
Db        264 -TDVLYKWQLVSSKLPQDLFLRAMP------KPVNGVV----PSEKTIAVVFYAQFLGSA 312

Qy        289 EREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDY 348
              | :|       |  :::  : |     :      |  :||       : || :    :|
Db        313 RRLMAIMNKNLPELGLKR--EDCYEMSWINTTTFWQNYPV------GTSTSVL----LDR 360

Qy        349 PFYALTETINGSGP-NQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQ 407
              |          |||      | || |: |  |  ::: |||  :                :  :  :|
Db        361 P----------SGPAGAFYKSKSDYVKKPIPKEEMEKIWKAM---------LKFNNMWMQ 401

Qy        408 VDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGG 466
              : :|| : |:  ||||    |:  : |:||    | : :    ||  :|| | || |||
Db        402 WNPYGGVMDKIPADATAFPHRKGNLFKIQYFALWTDANATYANLGLMRDIYHEM-EPY-- 458

Qy        467 VPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGK--------YGALELYFLGNLNRLIKA 518
                  | |         : || |:|: :  :|:         ||:   |||||  ||::
Db        459 -------VSSNP---REAFLNYRDIDVGSNPSGETNLEEAKIYGS--KYFLGNFKRLMEV 506

Qy        519 KWLWDPNEIFTNKQSIP 535
              |  :||    |  :||||
Db        507 KAKYDPENFFRFEQSIP 523


RESULT 14
Q9AYM8
ID   Q9AYM8      PRELIMINARY;      PRT;   535 AA.
AC   Q9AYM8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CPRD2 protein.
GN   CPRD2.
OS   Vigna unguiculata (Cowpea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX   NCBI_TaxID=3917;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Iuchi S.;
RT   "Drought inducible gene from cowpea.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB056448; BAB33033.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
SQ   SEQUENCE   535 AA;  60291 MW;  7F889FCBF52FD80E CRC64;

  Query Match             11.5%;  Score 346;  DB 10;  Length 535;
  Best Local Similarity   26.2%;  Pred. No. 2.9e-18;
```

```
  Matches  138;  Conservative   70;  Mismatches  207;  Indels  112;  Gaps   24;

Qy         49 VVYTPQGACTALDRAMEKCSPG---TVRIVSGGHCYEDFVFDECVK-AIINVTGLVESGY 104
              |: |||    :  :|  |||      :|  |||| ||   :    |   |:::    |
Db         78 VIVTPQ--VVSHIQATIKCSQRHGLQIRTRSGGHDYEGLSYVARVPFVILDLLNFREIKV 135

Qy        105 D-DDRGYFVSSGDTNWGSFKTLFRDHGRV-LPGGSCYSVGLGGHIVGGGDGILARLHGLP 162
              | ::|  :|  | |    : |: :    :  | | ||||| |||| ||| | | | :||
Db        136 DVENRTAWVQVGATLGELYYTISQASKTLGFPAGVCYSVGAGGHISGGGYGFLMRKYGLA 195

Qy        163 VDWLSGVEVVVKPVLTEDSVLKYVHKDSEGN-------DGELFWAHTGGGGGNFGIITKY 215
               | :    ::                |  ||         :||||  |||| :||:|  :
Db        196 ADNVIDAHII---------------DVNGNLLDRKAMGEDLFWAIRGGGGASFGVIVSW 239

Qy        216 YFKDLPMSPRGVI----------ASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGK 265
              | :|:    :           |: :   |     : | : :     |||  :  :  ||
Db        240 KIKLVPVPSTVTVFNVERILEENATEIIEKWQ-LVANKLDERIFLRMDLARAN-SSQHGK 297

Qy        266 FQI-------FHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALG 318
              :        |   ||  :  :  ::       |:   |                 | |  |  :|
Db        298 LALQANFVAMFQGGVEELIPLMQKNFP-----ELGLKR-----------KDCTETSWIG 340

Qy        319 GHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFP 378
                | |          :  : :      : |   |  |    ||    : | ||  |: |  |
Db        341 S----AVF---------TNGALIGSSGHEAPEVLLNRTQIRSG---KYKGKSDYVRKPIP 384

Qy        379 DFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQ 437
               :  :|::| :          ::: : ||    :||::  :          | |  ||  : |
Db        385 VDGLRGLWRWLND-------DKVQYSQLQFAPYGGKMDNISESEIPFAHRSGYIFHIHYV 437

Qy        438 TYWQEEDKDAV--NLKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNN 495
              ||||  :|   :: |||  |: | |||          | :         | || |:|:
Db        438 VVWQEEGDEATQRHVNWIRRLYKYM-EPY---------VSNSP---RAAYVNYRDLDIGV 484

Qy        496 WKNG--KYGALEL----YFLGNLNRLIKAKWLWDPNEIFTNKQSIPT 536
                 ||   |    :    ||  |  ||    |   ||:  | |:|||||
Db        485 NNNGYTSYHQASIWGLKYFSNNFKRLATVKTKVDPHNFFRNEQSIPT 531


RESULT 15
Q9FZC8
ID   Q9FZC8      PRELIMINARY;      PRT;   529 AA.
AC   Q9FZC8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   T1K7.20 protein.
GN   T1K7.20.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Sakano H., Liu S.X., Yu G., Lee J.M., Lenz C., Pham P., Toriumi M.,
RA   Chin C., Chiou J., Choi E., Chung M., Gonzalez A., Howng B., Liu A.,
RA   Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
RA   Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA   Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
RA   Ecker J.R., Federspiel N.A., Theologis A.;
RT   "The sequence of BAC T1K7 from Arabidopsis thaliana chromosome 1.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC013427; AAF98574.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
SQ   SEQUENCE   529 AA;  59217 MW;  0DC786AA3CB0590D CRC64;

  Query Match             11.1%;  Score 333.5;  DB 10;  Length 529;
  Best Local Similarity   25.8%;  Pred. No. 2.6e-17;
  Matches  128;  Conservative   77;  Mismatches  184;  Indels  107;  Gaps   22;

Qy         72 VRIVSGGHCYEDFVFDECVKAII-------NVTGLVESGYDDDRGYFVSSGDTNWGSFKT 124
              :|| ||||  |   :   |  :|        ::|  | |     :  :| :| |  |
Db        103 LRIRSGGHDLEGLSYRSSVPFVILDMFNLRSITVNVLS-----KKAWVQAGAT-LGELYV 156

Qy        125 LFRDHGRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSV 182
                 :  : |  | | | :||:|||| ||| | | |  |: || :|  :::       |
Db        157 KINEASQTLAFPAGVCPTVGVGGHISGGGYGNLMRKFGITVDHVSDAQLI-------DVN 209

Qy        183 LKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPR--GVIASNLHFSWDGFT 240
              | ::: | | |  ||||  |||| :||:|  :   :|   |:  |   |      |
Db        210 GKLLNRASMGED--LFWAIRGGGGASFGVILSWKI-NLVKVPKILTVFKVNKTLEQGG-- 264

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEEFV----------MYLYTSYSND 287
                |:| |         |:    ||  :| :|  : :           :  |  :
Db        265 ----TDVLYK--------WQLVATKFPEDLFMRAWPQIINGAERGDRTIAVVFYAQFLGP 312

Qy        288 AEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMD 347
              |:: :|        |  : :  : |             || :|    |       | : |   :
Db        313 ADKLLAIMNQRLPELGLRR--EDCHEMSWFNTTLFWADYPAGTPK------SVLLDRPTN 364

Qy        348 YPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDAL-L 406
               |:                | || |: |  |   :: :||          |  :  : : :
Db        365 PGFF--------------KSKSDYVKKPIPKEGLEKLWK---------TMFKFNNIVWM 400

Qy        407 QVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYG 465
              | : :|| : ::    |||    |:  : |:|| | |    :  ::|  ::: | :: |||
Db        401 QFNPYGGVMDQIPSTATAFPHRKGNMFKVQYSTTWLAANATEISLSMMKELY-KVAEPY- 458

Qy        466 GVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGALEL------YFLGNLNRLIKAK 519
                    | |          :||| |:|: :  : :    |       ||||||  ||:: |
Db        459 --------VSSNP---REAFFNYRDIDIGSNPSDETNVDEAKIYGYKYFLGNLKRLMQVK 507

Qy        520 WLWDPNEIFTNKQSIP 535
               :||   | |:||||
Db        508 AKYDPENFFKNEQSIP 523


Search completed: June 10, 2003, 09:59:59
Job time : 38 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 09:49:56 ; Search time 24 Seconds
(without alignments)
943.586 Million cell updates/sec

Title: US-09-998-284-2
Perfect score: 2997
Sequence: 1 MATLPQKDPGYIVIDVNAGT.....IFTNKQSIPTKPLKEPKQTK 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	290.5	9.7	535	1	RETO_PAPSO	P93479 papaver som
2	286.5	9.6	538	1	RETO_ESCCA	P30986 eschscholzi
3	154.5	5.2	447	1	MCRA_STRLA	P43485 streptomyce
4	118.5	4.0	769	1	DIVL_CAUCR	Q9rqq9 caulobacter
5	116	3.9	478	1	MM03_RABIT	P28863 oryctolagus
6	111	3.7	1009	1	YE68_METJA	Q58863 methanococc
7	109.5	3.7	496	1	DLD3_YEAST	P39976 saccharomyc
8	104	3.5	661	1	YJCS_ECOLI	P32717 escherichia
9	103	3.4	634	1	ELM1_ASPFU	P46074 aspergillus
10	102	3.4	634	1	ELM2_ASPFU	P46075 aspergillus
11	102	3.4	702	1	AMYG_CLOSP	P29761 clostridium
12	101.5	3.4	969	1	DPOM_NEUIN	P33538 neurospora
13	97.5	3.3	836	1	MCEL_SFVKA	P25950 shope fibro
14	97	3.2	469	1	MM01_BOVIN	P28053 bos taurus
15	97	3.2	955	1	AGAB_VIBS7	P48840 vibrio sp.
16	95.5	3.2	530	1	AIP2_YEAST	P46681 saccharomyc
17	95	3.2	847	1	AGUA_TRIRE	Q99024 trichoderma
18	94.5	3.2	359	1	RF2_BORBU	O51101 borrelia bu
19	94.5	3.2	494	1	K1CL_HUMAN	Q99456 homo sapien
20	94.5	3.2	1301	1	PTP9_DROME	P35832 drosophila
21	94	3.1	280	1	SPEE_AQUAE	O66473 aquifex aeo
22	94	3.1	388	1	CWF2_SCHPO	P87126 schizosacch
23	93.5	3.1	562	1	EST1_CAEEL	Q04457 caenorhabdi
24	92.5	3.1	290	1	HTRL_ECOLI	P25666 escherichia
25	92.5	3.1	450	1	DCOR_CHICK	P27118 gallus gall
26	92	3.1	1287	1	RPO1_FOWPV	Q9j593 fowlpox vir
27	92	3.1	4655	1	LRP2_HUMAN	P98164 homo sapien
28	91.5	3.1	397	1	PAPS_BACSU	P42977 bacillus su
29	91.5	3.1	458	1	HDNO_ARTOX	P08159 arthrobacte
30	91.5	3.1	663	1	MM02_CHICK	Q90611 gallus gall
31	91.5	3.1	1290	1	BXB_CLOBO	P10844 clostridium
32	91	3.0	787	1	K6PF_DROME	P52034 drosophila
33	90.5	3.0	477	1	MM03_MOUSE	P28862 mus musculu
34	90.5	3.0	534	1	FMO2_HUMAN	Q99518 homo sapien
35	90.5	3.0	559	1	HUTU_BACHD	Q9kbe5 bacillus ha
36	90.5	3.0	1263	1	RPOB_THEMA	P29398 thermotoga
37	90	3.0	477	1	NIFD_METMP	P71526 methanococc
38	90	3.0	946	1	K6P2_CANAL	O94200 candida alb
39	90	3.0	990	1	K6P1_PICPA	Q92448 pichia past
40	89.5	3.0	467	1	MM08_HUMAN	P22894 homo sapien
41	89.5	3.0	514	1	SYS_METMP	O30520 methanococc
42	89.5	3.0	1006	1	BGAL_ASPNG	P29853 aspergillus
43	89.5	3.0	2366	1	TOXB_CLODI	P18177 clostridium
44	89	3.0	680	1	GAOA_DACDE	Q01745 dactylium d
45	89	3.0	987	1	K6P1_YEAST	P16861 saccharomyc

ALIGNMENTS

RESULT 1
RETO_PAPSO
ID RETO_PAPSO STANDARD; PRT; 535 AA.
AC P93479;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Reticuline oxidase precursor (EC 1.5.3.9) (Berberine-bridge-forming
DE enzyme) (BBE) (Tetrahydroprotoberberine synthase).
GN BBE1.
OS Papaver somniferum (Opium poppy).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
OC Papaveraceae; Papaver.
OX NCBI_TaxID=3469;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Marianne;
RX MEDLINE=97127853; PubMed=8972604;
RA Facchini P.J., Penzes C., Johnson A.G., Bull D.;
RT "Molecular characterization of berberine bridge enzyme genes from
RT opium poppy.";
RL Plant Physiol. 112:1669-1677(1996).
CC -!- FUNCTION: Essential to the formation of benzophenanthridine
CC alkaloids in the response of plants to pathogenic attack.
CC Catalyzes the stereospecific conversion of the N-methyl moiety of
CC (S)-reticuline into the berberine bridge carbon of (S)-scoulerine.
CC -!- CATALYTIC ACTIVITY: (S)-reticuline + O(2) = (S)-scoulerine +
CC H(2)O(2).
CC -!- COFACTOR: FAD AND METAL ION.
CC -!- PATHWAY: Benzophenanthridine alkaloids biosynthesis.
CC -!- SUBCELLULAR LOCATION: VESICULAR.
CC -!- SIMILARITY: Belongs to the oxygen-dependent FAD-linked
CC oxidoreductase family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AF025430; AAC61839.1; -.
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
DR PROSITE; PS00862; OX2_COVAL_FAD; 1.
KW Oxidoreductase; Signal; Glycoprotein; Flavoprotein; FAD;
KW Alkaloid metabolism.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 535 RETICULINE OXIDASE.
FT BINDING 108 108 FAD (COVALENT) (BY SIMILARITY).
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 535 AA; 59903 MW; F0341EF38AB41239 CRC64;

```
  Query Match             9.7%;  Score 290.5;  DB 1;  Length 535;
  Best Local Similarity   23.8%;  Pred. No. 7.5e-15;
  Matches  123;  Conservative   87;  Mismatches  160;  Indels  147;  Gaps   28;

Qy         71 TVRIVSGGHCYE--------DFVFDECVKA------IINVTGLVESGYDDDRGYFVSSGD 116
              |:|: |||| ||          ||  : :         ::: |  ||||       |:  :
Db        100 TIRLRSGGHSYEGLSYTADTPFVIVDMMNLNRISIDVLSETAWVESGATLGELYYAIAQS 159

Qy        117 TNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPV 176
              |:   |              | | :|| |||| ||| |:::| :||  |        ||  :
Db        160 TDTLGF-----------TAGWCPTVGSGGHISGGGFGMMSRKYGLAAD------NVVDAI 202

Qy        177 LTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVI-------- 228
              |  ||    : ::  |:|  :|||  ||||| :| |  :  | ||:  :  :
Db        203 LI-DSNGAILDREKMGDD--VFWAIRGGGGGVWGAIYAWKIKLLPVPEKLTVFRVTKNVG 259

Qy        229 ---ASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYS 285
                ||:|   |  :  | | :  |           : :|         : ::|:|
Db        260 IEDASSLLHKWQ-YVADELDEDFT----------VSVLGGV----NGNDAWLMFLGLHLG 304

Qy        286 -NDAEREVAQDRHYHL---EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYM 341
               || : :  ::   |   : : :::                |
Db        305 RKDAAKTIIDEKFPELGLVDKEFQEM-------------SWG---------------- 333

Qy        342 HDETMDYPFYALTETINGSGPNQRGKY-KSAYMIK-DFPDFQIDV-IWKYLTEVPDGLTS 398
               |:|   | :  :||:     |:  |: : |:  | ||    : : ::::  |:
Db        334 --ESM--AFLSGLDTISELN-NRFLKFDERAFKTKVDFTKVSVPLNVFRHALEM-----L 383

Qy        399 AEMKDALLQVDMFGGEIHKVVWDATAVAQR-------EYIIKLQYQTYWQEEDKDAVN-- 449
              :|     : :: |||:: ::  | |      |        ||||        | ::::  :
Db        384 SEQPGGFIALNGFGGKMSEISTDFTPFPHRKGTKLMFEYIIA------WNQDEESKIGEF 437

Qy        450 LKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLN--NWKN--GKYGALE 505
              :|:  ||: :         :|     |    |      | |: |:|:   :|:|      |:|
Db        438 SEWLAKFYDYL--------EPFVSKEPRVG-----YVNHIDLDIGGIDWRNKSSTTNAVE 484

Qy        506 L-------YFLGNLNRLIKAKWLWDPNEIFTNKQSIP 535
              :       ||  |  ||:||| | ||| :| : ||||
Db        485 IARNWGERYFSSNYERLVKAKTLIDPNNVFNHPQSIP 521


RESULT 2
RETO_ESCCA
ID   RETO_ESCCA     STANDARD;      PRT;   538 AA.
AC   P30986;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Reticuline oxidase precursor (EC 1.5.3.9) (Berberine-bridge-forming
DE   enzyme) (BBE) (Tetrahydroprotoberberine synthase).
GN   BBE1.
OS   Eschscholzia californica (California poppy).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
OC   Papaveraceae; Eschscholzioideae; Eschscholzia.
OX   NCBI_TaxID=3467;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92052284; PubMed=1946465;
RA   Dittrich H., Kutchan T.M.;
RT   "Molecular cloning, expression, and induction of berberine bridge
RT   enzyme, an enzyme essential to the formation of benzophenanthridine
RT   alkaloids in the response of plants to pathogenic attack.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:9969-9973(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98145481; PubMed=9484487;
RA   Hauschild K., Pauli H.H., Kutchan T.M.;
RT   "Isolation and analysis of a gene bbe1 encoding the berberine bridge
RT   enzyme from the California poppy Eschscholzia californica.";
RL   Plant Mol. Biol. 36:473-478(1998).
CC   -!- FUNCTION: Essential to the formation of benzophenanthridine
CC       alkaloids in the response of plants to pathogenic attack.
CC       Catalyzes the stereospecific conversion of the N-methyl moiety of
CC       (S)-reticuline into the berberine bridge carbon of (S)-scoulerine.
CC   -!- CATALYTIC ACTIVITY: (S)-reticuline + O(2) = (S)-scoulerine +
CC       H(2)O(2).
CC   -!- COFACTOR: FAD AND METAL ION.
CC   -!- PATHWAY: Benzophenanthridine alkaloids biosynthesis.
CC   -!- SUBCELLULAR LOCATION: VESICULAR.
CC   -!- SIMILARITY: Belongs to the oxygen-dependent FAD-linked
CC       oxidoreductase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S65550; AAB20352.1; -.
DR   EMBL; AF005655; AAC39358.1; -.
DR   PIR; A41533; A41533.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
DR   PROSITE; PS00862; OX2_COVAL_FAD; 1.
KW   Oxidoreductase; Signal; Glycoprotein; Flavoprotein; FAD;
KW   Alkaloid metabolism.
FT   SIGNAL        1     23
FT   CHAIN        24    538       RETICULINE OXIDASE.
FT   BINDING     104    104       FAD (COVALENT) (BY SIMILARITY).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   538 AA;  59958 MW;  1A505F86A06CDB24 CRC64;

  Query Match             9.6%;  Score 286.5;  DB 1;  Length 538;
  Best Local Similarity   24.8%;  Pred. No. 1.5e-14;
  Matches  128;  Conservative   72;  Mismatches  170;  Indels  147;  Gaps   27;

Qy         71 TVRIVSGGHCYEDFVF-DECVKAIINVTGLVESGYD-DDRGYFVSSGDTNWGSFKTLFRD 128
              |:|: |||| ||    :  :     :|::  |      | :     :| || |     :  :
Db         96 TIRLRSGGHSYEGLSYTSDTPFILIDLMNLNRVSIDLESETAWVESGSTLGELYYAITES 155

Qy        129 HGRV-LPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTE--DSVLKY 185
               ::    | | :|| |||| ||| |:::| :||  |        ||  :| :   ::|
Db        156 SSKLGFTAGWCPTVGTGGHISGGGFGMMSRKYGLAAD------NVVDAILIDANGAIL-- 207

Qy        186 VHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVI----------ASNLHF 234
                | :    ::|||  ||||| :| |  :  | ||:  :  :           |::|
Db        208 ---DRQAMGEDVFWAIRGGGGGVWGAIYAWKIKLLPVPEKVTVFRVTKNVAIDEATSLLH 264

Qy        235 SWDGFTRDALQDLLTKYFKLARCD----WKNTVG-KFQIFHQAAEEFVM---------- 278
               |  |  : |::  |    |   |    |   :|  | :   |  | :
Db        265 KWQ-FVAEELEEDFT-LSVLGGADEKQVWLTMLGFHFGLKTVAKSTFDLLFPELGLVEED 322

Qy        279 YLYTSYSND-----AEREVAQDRHYHLEADIEQIYKT-CEPTKALGGHAGWAPFPVRPRK 332
               ||  |:             |:|  :  |: | |: :||  : ||          | | :
Db        323 YLEMSWGESFAYLAGLETVSQLNNRFLKFD-ERAPKTKVDLTK--------EPLPSK--- 370

Qy        333 RHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEV 392
                              || | |                          |::
Db        371 ---------------AFYGLLER--------------------------------LSKE 382

Qy        393 PDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYI-IKLQYQTYW--QEEDKDAVN 449
              |:|          : :: |||:: |:  | |     |    : ::|   |   |: |
Db        383 PNG---------FIALNGFGGQMSKISSDFTPFPHRSGTRLMVEYIVAWNQSEQKKKTEF 433

Qy        450 LKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLN--NWKNGKY--GALE 505
              | |:   || | :|:      | ::          | |: |:||   :| |      |:|
Db        434 LDWLEKVYEFM-KPF---VSKNPRLG---------YVNHIDLDLGGIDWGNKTVVNNAIE 480

Qy        506 L-------YFLGNLNRLIKAKWLWDPNEIFTNKQSIP 535
              :       ||| |  |||:|| | ||| :| : ||||
Db        481 ISRSWGESYFLSNYERLIRAKTLIDPNNVFNHPQSIP 517
```

```
RESULT 3
MCRA_STRLA
ID   MCRA_STRLA     STANDARD;      PRT;   447 AA.
AC   P43485;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Mitomycin radical oxidase (EC 1.5.3.-).
GN   MCRA.
OS   Streptomyces lavendulae.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1914;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC   STRAIN=B619;
RX   MEDLINE=94292477; PubMed=7517396;
RA   August P.R., Flickinger M.C., Sherman D.H.;
RT   "Cloning and analysis of a locus (mcr) involved in mitomycin C
RT   resistance in Streptomyces lavendulae.";
RL   J. Bacteriol. 176:4448-4454(1994).
CC   -!- FUNCTION: INVOLVED IN MITOMYCIN RESISTANCE; OXIDIZES REDUCED FORM
CC       OF MITOMYCINS.
CC   -!- COFACTOR: FAD (POTENTIAL).
CC   -!- INDUCTION: BY MITOMYCIN C IN CONCENTRATIONS AS LOW AS 300 NM.
CC   -!- SIMILARITY: Belongs to the oxygen-dependent FAD-linked
CC       oxidoreductase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L29247; AAA21476.1; -.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
DR   PROSITE; PS00862; OX2_COVAL_FAD; 1.
KW   Antibiotic resistance; Oxidoreductase; Flavoprotein; FAD.
FT   INIT_MET      0      0
FT   BINDING      63     63       FAD (COVALENT) (BY SIMILARITY).
SQ   SEQUENCE   447 AA;  48220 MW;  3B0071BB818D838C CRC64;

  Query Match             5.2%;  Score 154.5;  DB 1;  Length 447;
  Best Local Similarity   22.4%;  Pred. No. 0.00018;
  Matches   85;  Conservative   51;  Mismatches  163;  Indels   81;  Gaps   17;

Qy         21 PDKP---DPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTALDRAMEKCSPGTVRIVSG 77
              ||:|   | ||      | ||    :   :|      |    |:  | |:  |  | :::
Db         10 PDQPGYDDARL-----GLNRA-AESRPAYVVEAADEQEVAAAVRLAAEQKRP--VGVMAT 61

Qy         78 GHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGS 137
              ||      |: |  ::|   :     |  |          |          ||     ||
Db         62 GH-GPSVSADDAV--LVNTRRMEGVSVDAARATAWIEAGARWRKVLEHTAPHGLAPLNGS 118

Qy        138 CYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGEL 197
               :||  |::|||| |:| |  |   | :  : :|     | |  |: |   : | | :|
Db        119 SPNVGAVGYLVGGGAGLLGRRFGYAADHVRRLRLV-----TADGRLRDV---TAGTDPDL 170

Qy        198 FWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARC 257
              |||   ||  |||::       |::   :    |:|:      :|  ::|  | :  |
Db        171 FWA-VRGGKDNFGLVVGMEVDLFPVT--RLYGGGLYFA-----GEATAEVLHAYAEWVRH 222

Qy        258 DWKNTVGKFQIFHQ---------AAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIY 308
               :    : |              |: :|  :|| :           || | :
Db        223 VPEEMASSVLLVHNPDLPDVPEPLRGRFITHLRIAYSGE------------PADGEHLV 269

Qy        309 KTCEPTKALGGHAGWAPFPV---RPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQR 365
              :   | : ||        |:     |    ::   :| |    |: |

Db        270 R---PLRELG--------PILLDTVRDMPYAEVGTIHHEPTSMPYVA------------ 305

Qy        366 GKYKSAYMIKDFPDFQIDVI 385
                 |    :: |  |  :|:|
Db        306 --YDRNVLLSDLTDDAVDII 323


RESULT 4
DIVL_CAUCR
ID   DIVL_CAUCR     STANDARD;      PRT;   769 AA.
AC   Q9RQQ9; Q9A2S2;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sensor protein divL (EC 2.7.3.-).
GN   DIVL OR CC3484.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A., AND AUTOPHOSPHORYLATION SITE.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=20027501; PubMed=10557274;
RA   Wu J., Ohta N., Zhao J.L., Newton A.;
RT   "A novel bacterial tyrosine kinase essential for cell division and
RT   differentiation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:13068-13073(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC   -!- FUNCTION: Required for cell division and growth It catalyzes the
CC       phosphorylation of CtrA and activates transcription in vitro of
CC       the cell cycle-regulated fliF promoter.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: Autophosphorylated.
CC   -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF083422; AAF08344.2; -.
DR   EMBL; AE006007; AAK25446.1; -.
DR   TIGR; CC3484; -.
DR   PhosSite; Q9RQQ9; -.
DR   InterPro; IPR004358; Bact_sens_pr_C.
DR   InterPro; IPR003661; His_kinA.
DR   InterPro; IPR004359; HIS_KIN_sig.
DR   InterPro; IPR000014; PAS_domain.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00512; signal; 1.
DR   PRINTS; PR00344; BCTRLSENSOR.
DR   SMART; SM00387; HATPase_c; 1.
DR   SMART; SM00388; HisKA; 1.
DR   TIGRFAMs; TIGR00229; sensory_box; 2.
DR   PROSITE; PS50109; HIS_KIN; 1.
KW   Sensory transduction; Transferase; Kinase; Transmembrane;
KW   Phosphorylation; Complete proteome.
```

```
FT   TRANSMEM      6     26       POTENTIAL.
FT   DOMAIN      547    758       HISTIDINE KINASE.
FT   DOMAIN        9    221       ALA-RICH.
FT   MOD_RES     550    550       PHOSPHORYLATION (AUTO-).
FT   CONFLICT    200    200       Q -> H (IN REF. 1).
FT   CONFLICT    216    216       V -> E (IN REF. 1).
SQ   SEQUENCE   769 AA;  82796 MW;  002B2428F18A57EF CRC64;

  Query Match             4.0%;  Score 118.5;  DB 1;  Length 769;
  Best Local Similarity   22.4%;  Pred. No. 0.23;
  Matches  109;  Conservative   69;  Mismatches  192;  Indels  117;  Gaps   30;

Qy         59 ALDRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAI---INVTGLVESGYDDDRGYFVSSG 115
              | | |:     |   :|:||   |  :   | ||:     |: :| :         :|
Db         58 AFDTAVIAVEGGRANLVAGG---EGLI--ACAKALGADAEVSAVVAA---------LSDA 103

Qy        116 DTNWGSFKTLFRDHGRVLPGGSCYSVGLGGH----IVGGGDGILARLHGLPVDWL-SGVE 170
              | |:    |   : |       |     | |    : |   | ||| |   |:|   ||:
Db        104 DPNYAQKLTALFERGE-----PCVFEARGPHGLVSVEGRAAGALAWLRLAPIDRADSGLP 158

Qy        171 VVVKPVLTEDSVLK--YVHKDSEGNDGELFWAH------TGGGGGNFGIITKYYFKDLPM 222
                :   |||::  ::     | ||:  | :       |       :    |
Db        159 TAARFAAFVDSVVEPCWI----AGADGQAIWGNAAFVRAVGAASAQAPALAGKSF----- 209

Qy        223 SPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEF---VMY 279
              ||  |  :  :  |  |:||:           | |  |: : |  :|:     :
Db        210 -DRGADAVVVEAAGKGERREALR-----------WINVEGRRRAFRLSAQPLDGGGVG 255

Qy        280 LYTSYSNDAEREVAQDRHYHLEADIEQIYKTCE------PTKALGGHAG-----WAPFPV 328
              :: :   : | :|      |:||  | :    |       |: |  |        |   |
Db        256 VFCADVTEIE-DVRDAFKKHVEAHDETLNHIAEAVAIFSQTRRLSYHNTAFAELWGLEPA 314

Qy        329 RPRKRHTSKTSYMHDETMD--YPFYALTETINGSGPNQRGKYKSAYMIKDFPDF--QIDV 384
                | |      | | :|       | |||: :|      :|:| : : :  |   | |
Db        315 WLADRPT------HGEVLDRLRQRRRLPETIDYAG------WKAAELAR-YEDLGPQADD 361

Qy        385 IWKYLTEVPDGLTSAEMKDALLQVDMFGGEI---HKVVWDATAVAQREYIIKLQYQTYWQ 441
              :|    ::||| |   ::    |    || :     :  :    ||   :|::|  | :
Db        362 LW----DLPDGRTLKVVR----QPHPLGGMLLIYSDITGELRLKAQYNALIQVQQATLDK 413

Qy        442 EEDKDAV-----NLKWIRDFYEEMYEPYGGVPDPNTQVESG--KGVFEGCYFNYPDV-DL 493
                | ||      |:     : | :| :  |  |:    :|  :|| | |    | : ||
Db        414 LNDAVAVFGSDGRLR----LHNEAFETFWNV-TPHALEAAGDFEGVVELC---VPRLHDL 465

Qy        494 NNWKNGK 500
              : |:  |
Db        466 SFWRELK 472


RESULT 5
MM03_RABIT
ID   MM03_RABIT     STANDARD;      PRT;   478 AA.
AC   P28863;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Stromelysin-1 precursor (EC 3.4.24.17) (Matrix metalloproteinase-3)
DE   (MMP-3) (Transin-1) (SL-1).
GN   MMP3.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88077214; PubMed=2825726;
RA   Fini M.E., Karmilowicz M.J., Ruby P.L., Beeman A.M., Borges K.A.,
RA   Brinckerhoff C.E.;
RT   "Cloning of a complementary DNA for rabbit proactivator. A
RT   metalloproteinase that activates synovial cell collagenase, shares
RT   homology with stromelysin and transin, and is coordinately regulated
RT   with collagenase.";
RL   Arthritis Rheum. 30:1254-1264(1987).
RN   [2]
RP   SEQUENCE OF 1-167 FROM N.A.
RX   MEDLINE=87156645; PubMed=3030290;
RA   Whitham S.E., Murphy G., Angel P., Rahmsdorf H.J., Smith B.,
RA   Lyons A., Harris T.J.R., Reynolds J.J., Herrlich P., Docherty A.J.P.;
RT   "Comparison of human stromelysin and collagenase by cloning and
RT   sequence analysis.";
RL   Biochem. J. 240:913-916(1986).
CC   -!- FUNCTION: CAN DEGRADE FIBRONECTIN, LAMININ, GELATINS OF TYPE I,
CC       III, IV, AND V; COLLAGENS III, IV, X, AND IX, AND CARTILAGE
CC       PROTEOGLYCANS. ACTIVATES PROCOLLAGENASE.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage where P1', P2' and P3'
CC       are hydrophobic residues.
CC   -!- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.
CC   -!- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M25664; AAA31467.1; -.
DR   PIR; A37306; KCRBS1.
DR   HSSP; P08254; 1SLM.
DR   MEROPS; M10.005; -.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR001818; Matrixin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF00045; hemopexin; 4.
DR   Pfam; PF00413; Peptidase_M10; 1.
DR   PRINTS; PR00138; MATRIXIN.
DR   SMART; SM00120; HX; 4.
DR   SMART; SM00235; ZnMc; 1.
DR   PROSITE; PS00024; HEMOPEXIN; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00546; CYSTEINE_SWITCH; 1.
KW   Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;
KW   Collagen degradation; Extracellular matrix; Signal.
FT   SIGNAL        1     17       PROBABLE.
FT   PROPEP       18    100       ACTIVATION PEPTIDE.
FT   CHAIN       101    478       STROMELYSIN-1.
FT   DOMAIN      288    478       HEMOPEXIN-LIKE.
FT   SITE         93     93       CYSTEINE SWITCH (POTENTIAL).
FT   METAL       219    219       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    220    220       BY SIMILARITY.
FT   METAL       223    223       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   CARBOHYD    121    121       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    291    478       BY SIMILARITY.
FT   CONFLICT     83     83       N -> D (IN REF. 2).
FT   CONFLICT    128    128       R -> K (IN REF. 2).
SQ   SEQUENCE   478 AA;  53942 MW;  CA742E31A4549D40 CRC64;

  Query Match             3.9%;  Score 116;  DB 1;  Length 478;
  Best Local Similarity   20.7%;  Pred. No. 0.19;
  Matches  101;  Conservative   59;  Mismatches  145;  Indels  182;  Gaps   24;

Qy         38 RWIGTNIDFVYVVYTPQGACTALDRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVT 97
              :|  |:: :  | |||     |:| |:||                      :|    ||
Db        109 KWTKTHLTYRIVNYTPDLPRDAVDAAIEK---------------------ALKVWEEVT 146

Qy         98 GLVES-GYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILA 156
               |  |  |: :    :| |          |:||  :|
Db        147 PLTFSRKYEGEADIMISFG----------VREHGDFIP---------------------- 174

Qy        157 RLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYY 216
                                          |  ||   |  |:  | |
Db        175 ------------------------------FDGPGN--VLAHAYAPGPG---------- 191
```

```
Qy        217 FKDLPMSPRGVIASNLHFSWD-GFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEE 275
Db        192 -----------INGDAHFDDDEQWTKDTTG---TNLFLVAAHELGHSLG---LFHSANPE 234

Qy        276 FVMY-LYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRH 334
Db        235 ALMYPVYNAFTDLARFRLSQD-------DVDGIQSLYGPAPASPDNSGVPMEPVPP---- 283

Qy        335 TSKTSYMHDETMDYPFYALTETINGSGPNQRGKY---KSAYMIKDFPDFQ-IDVIWKYLT 390
Db        284 GSGTPVMCDP--DLSFDAIS-TLRGEILFFKDRYFWRKSLRILE--PEFHLISSFWPSLP 338

Qy        391 EVPDGLTSAEMKDALLQVDMFGG-----------------EIHKVVWDAT------AVAQ 427
Db        339 SAVDAAYEVISRDT---VFIFKGTQFWAIRGNEVQAGYPRSIHTLGFPSTIRKIDAAISD 395

Qy        428 REYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEP------YGGVPDPNTQVESGKGVF 481
Db        396 KE-----RKKTYFFVEDK-----YWRFDEKRQSLEPGFPRHIAEDFPGINPKIDA---VF 442

Qy        482 EGCYFNY 488
Db        443 EAFGFFY 449


RESULT 6
YE68_METJA
ID   YE68_METJA     STANDARD;      PRT;  1009 AA.
AC   Q58863;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein MJ1468.
GN   MJ1468.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: CONTAINS 5 PKD DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67588; AAB99478.1; -.
DR   TIGR; MJ1468; -.
DR   InterPro; IPR000601; PKD_domain.
DR   Pfam; PF00801; PKD; 6.
DR   SMART; SM00089; PKD; 7.
DR   PROSITE; PS50093; PKD; 5.
KW   Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT   TRANSMEM      6     26       POTENTIAL.
FT   TRANSMEM    985   1005       POTENTIAL.
FT   DOMAIN      213    247       PKD 1.
FT   DOMAIN      436    503       PKD 2.
FT   DOMAIN      724    806       PKD 3.
FT   DOMAIN      822    886       PKD 4.
FT   DOMAIN      925    962       PKD 5.
FT   DOMAIN      293    298       POLY-ASN.
SQ   SEQUENCE   1009 AA;  115119 MW;  13E9B4933EAB7972 CRC64;

  Query Match             3.7%;  Score 111;  DB 1;  Length 1009;
  Best Local Similarity   18.3%;  Pred. No. 1.2;
  Matches  113;  Conservative   70;  Mismatches  214;  Indels  222;  Gaps   26;

Qy         69 PGTVRIVSGGHCY-----------------EDFVFDECVKAIINVTGLVESGYDDDRGYF 111
Db         28 PNTMNITNSSGVYNTDPNIIAYNVNDTITFEALAPDSVAQDILD-NGVVKWDFGD----- 81

Qy        112 VSSGDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEV 171
Db         82 --LTETDYGNYRTTTHTYTFPFP----YPVAWCGYLNNTGYS-----KALTYNWL----- 125

Qy        172 VVKPVLTEDSVLKYVHKDSEGN---DGELFWAHTGGGGGNFGIITKYY--------FKDL 220
Db        126 ----VVGDVANTKYVFNGSPLNSKTSWEVYYNST-----NNTVIIKYYSETPVDREFNGL 176

Qy        221 PMSPRGVIASNLHFSWDGFTRDAL--QDLLTKYFKLAR----CDWKNTVGKFQIFHQAAE 274
Db        177 SVDTTSVTVN--------VSRDEIVEGDTVKFNFSVSRNIIFCVWSFGDGTFSF-----E 223

Qy        275 EFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRH 334
Db        224 KFPEHTYTKSGLYYPRVLVVDDSGRVM--VGYLDEGIEVKRARGGYIYWVTGP------- 274

Qy        335 TSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIW-KYL---- 389
Db        275 ----SHYDGEAYTYVYNSSGDDNNNNGNAYTDPYKITYKVNDTIKFEMSGAWGEYWKWDF 330

Qy        390 ---TEVPDGLTSAEMKDALLQVDMFGGEIHK---------------------------V 418
Db        331 GDGTETPYTYKS-----------YFTPSYHQYKFPFMWPFFWMSYGWGSWWKSDTLNFIV 379

Qy        419 VWDATAVAQREYIIKLQYQTYWQEE---DKDAVNLKW-------------IRDFY----- 457
Db        380 VDDVENTRYNFYPSSAHDKTYYDYEYNKENHTVNLYYYSDVISTPKFNVKLRDGYYIDIT 439

Qy        458 ------------------------------------------------EEMYEPYGGVPD 469
Db        440 ATADKTQVSVNENVRFDCSPYGNPIFIMWCFGDGTCSFEKSPTHRYSSSGLYYPHVFVID 499

Qy        470 PNTQVESG--KGVFEGCYFNYPDVDLNNWKNGKYGALEL---YFLGNLNRLIKAKWLWDP 524
Db        500 DNGNIEVGIPPPIGVGGYSSYPQI---------YASPTIAPTYYPINITIVEPASWTWYW 550

Qy        525 NEI-FTNKQSIPTKPLKEP 542
Db        551 HHIYFGDGGSVWIKPKKSP 569


RESULT 7
DLD3_YEAST
ID   DLD3_YEAST     STANDARD;      PRT;   496 AA.
AC   P39976;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable D-lactate dehydrogenase [cytochrome] (EC 1.1.2.4) (D-lactate
DE   ferricytochrome C oxidoreductase) (D-LCR).
GN   DLD3 OR YEL071W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=S288c / AB972;
RA   Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA   Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA   Chung E., Duncan M., Guzman E., Hartzell G., Hunicke-Smith S.,
RA   Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA   Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA   Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA   Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: (R)-lactate + 2 ferricytochrome c = pyruvate +
CC       2 ferrocytochrome c.
CC   -!- COFACTOR: FAD (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE FAD-BINDING OXIDOREDUCTASE/TRANSFERASE
CC       FAMILY 4.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U18795; AAB65016.1; -.
DR   SGD; S0000797; DLD3.
DR   InterPro; IPR004113; FAD-oxidase_C.
DR   InterPro; IPR001575; Oxid_FAD_bind.
DR   Pfam; PF01565; FAD_binding_4; 1.
DR   Pfam; PF02913; FAD-oxidase_C; 1.
KW   Oxidoreductase; Flavoprotein; FAD.
SQ   SEQUENCE   496 AA;  55225 MW;  4809F74EDF07F520 CRC64;

  Query Match             3.7%;  Score 109.5;  DB 1;  Length 496;
  Best Local Similarity   18.0%;  Pred. No. 0.62;
  Matches  110;  Conservative   82;  Mismatches  191;  Indels  227;  Gaps   31;

Qy          6 QKDPGYIVID---------------VNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYV 49
              :::| : |:|                 :|:  |::    | |  | : ::: | :     :
Db         19 KRNPNFKVLDSEDLAYFRSILSNDEILNSQAPEE----LASFNQDWMKKYRGQS----NL 70

Qy         50 VYTPQGACTALDRAMEKCSPGTVRIV-SGGHCYEDF------VFDECVKAIINVTGLVES 102
              :  |  :   : : |: |:   : :|  ||:   |       ||||  | :: |:  : :
Db         71 ILLP-NSTDKVSKIMKYCNDKKLAVVPQGGN--TDLVGASVPVFDEIVLSLRNMNKVRD- 126

Qy        103 GYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLP-----GGSCYSVGLGGHIVGGGDGI--- 154
               :|   | |                 ||  : |       :|    :|| :     |:
Db        127 -FDPVSGTFKCDAGVVMRDAHQFLHDHDHIFPLDLPSRNNCQ---VGGVVSTNAGGLNFL 182

Qy        155 -LARLHGLPVDWLSGVEVVVK--PVLTEDSVLKYVHKDSEGND-GELFWAHTGGGGGNFG 210
                |||     : |:|||:    :::  : |:   ||: | |  :||     |  |  |
Db        183 RYGSLHGN----VLGLEVVLPNGEIISNINALR---KDNTGYDLKQLFI----GAEGTIG 231

Qy        211 IITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFH 270
              ::|      :  : :    : : |  : |       | : | |  |: :    :  |:
Db        232 VVTGV---SIVAAAKPKALNAVFFGIENF------DTVQKLFVKAKSELSEILSAFEFMD 282

Qy        271 QAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRP 330
              : :                       :|  || :      | |      |||:
Db        283 RGS-------------------------IECTIEYL-------KDL-------PFPLEN 302

Qy        331 RKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGK----------------YKSAYMI 374
              :                 : || | ||   || |:|                      |
Db        303 Q----------------HNFYVLIET---SGSNKRHDDEKLTAFLKDTTDSKLISEGMMA 343

Qy        375 KDFPDFQIDVIWKYLTEVPDGL-------------------TSAEMKDALLQVDMFG-- 412
              ||  ||  | :| :   ||                     || : : |    : |
Db        344 KDKADF--DRLWTWRKSVPTACNSYGGMYKYDMSLQLKDLYSVSAAVTERLNAAGLIGDA 401

Qy        413 ---------------GEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFY 457
                           | ||        :| ||:                  | | |
Db        402 PKPVVKSCGYGHVGDGNIH------LNIAVREF-----------------TKQIEDLL 436

Qy        458 EEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIK 517
```

```
                |     |      :   | | |  :    :|   |          :|: |: ::
Db        437 EPFVYEYIASKKGSISAEHGIGFHKKGKLHYTRSD----------IEIRFMKDI----- 480

Qy        518 AKWLWDPNEI 527
                | :||| |
Db        481 -KNHYDPNGI 489


RESULT 8
YJCS_ECOLI
ID   YJCS_ECOLI     STANDARD;      PRT;   661 AA.
AC   P32717;
DT   01-OCT-1993 (Rel. 27, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein yjcS.
GN   YJCS OR B4083.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94089392; PubMed=8265357;
RA   Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA   Daniels D.L.;
RT   "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT   region from 89.2 to 92.8 minutes.";
RL   Nucleic Acids Res. 21:5408-5417(1993).
RN   [2]
RP   REVISION TO 614.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- SIMILARITY: TO PSEUDOMONAS SP. ALKYL SULFATASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00006; AAC43177.1; ALT_INIT.
DR   EMBL; AE000482; AAC77044.1; ALT_INIT.
DR   EcoGene; EG11955; yjcS.
DR   InterPro; IPR001279; Blactmase-like.
DR   Pfam; PF00753; lactamase_B; 1.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT    614    614       H -> D (IN REF. 1).
SQ   SEQUENCE   661 AA;  73150 MW;  64D5037ACEF5E455 CRC64;

  Query Match             3.5%;  Score 104;  DB 1;  Length 661;
  Best Local Similarity   20.9%;  Pred. No. 2.4;
  Matches  117;  Conservative   73;  Mismatches  208;  Indels  162;  Gaps   28;

Qy         15 DVNAGTPDKPDPRL--PSMKQGFNRRWIGT-----------NIDFVY---------VVY 51
              |:||  |:  :| |   |   | :  :  |             || ||           :
Db         93 DINAAAPETVNPSLWRQSQINGISGLFKVTDKMYQVRGQDISNITFVEGEKGIIVIDPLV 152

Qy         52 TPQGACTALDRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIIN--------VTGLVESG 103
              ||  |  |||   :      :  |   | : |      || ||:         |  :  :|
Db        153 TPPAAKAALDLYFQHRPQKPIVAVIYTHSHTDHYGG--VKGIISEADVKSGKVQVIAPAG 210

Qy        104 YDDDR-GYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSV--GLGGHIVGGGDGILAR--- 157
              : |:     | :|  |  | : |:  :| :||  :  :|  |||  :  |   |:|
```

```
Db        211 FMDEAISENVLAG--NIMSRRALY-SYGLLLPHNAQGNVGNGLGVTLATGDPSIIAPTKT 267

Qy        158 ---------LHGLPVDWL--SGVEV---------VVKPVLTEDSVLKYVHK--------- 188
                      : ||  |:|   | |           :| : | ::    :|
Db        268 IVRTGEKMIIDGLEFDFLMTPGSEAPAEMHFYIPALKALCTAENATHTLHNFYTLRGAKT 327

Qy        189 --------------DSEGNDGE-LFWAHTGGGGGNFGI---ITKY-----YFKDLPMSPR 225
                           |  ||| | ||  ||    ||  |   | ||     |  |
Db        328 RDTSKWTEYLNETLDMWGNDAEVLFMPHTWPVWGNKHINDYIGKYRDTIKYIHD------ 381

Qy        226 GVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYS 285
                 || :  |:| : : |::     ||  :| :       : | |   :  ||
Db        382 ----QTLHLANQGYTMNEIGDMIKLPPALAN-NWASRGYYGSVSHNARAVYNFYLGYYDG 436

Qy        286 NDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHA---------------GWAP----- 325
                | |      : | : : :: : |      :|||| |                 |:
Db        437 NPA------NLHPYGQVEMGKRY-----VQALGGSARVINLAQEANKQGDYRWSAELLKQ 485

Qy        326 -FPVRPRKRHTSKTSYMHDETMDYP--------FYALTETINGSGPNQRGKYKSAYMIKD 376
                  |  :       : | : |          ||     : |:   : | :| ::
Db        486 VIAANPGDQVAKNLQANNFEQLGYQAESATWRGFY-----LTGAKELREGVHKFSHGTTG 540

Qy        377 FPD----FQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYII 432
                ||      :::: ::    |   :|  |:  |  :|  |:   :  : : :  |:
Db        541 SPDTIRGMSVEMLFDFMAVRLDSAKAAG-KNISLNFNMSNGDNLNLTLNDSVLNYRK--- 596

Qy        433 KLQYQT----YWQEEDKDAV 448
                || |     |  ||  ||
Db        597 TLQPQADASFYISREDLHAV 616
```

RESULT 9
ELM1_ASPFU
ID ELM1_ASPFU STANDARD; PRT; 634 AA.
AC P46074;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Extracellular elastinolytic metalloproteinase precursor (EC 3.4.24.-).
GN MEP.
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 530-542.
RC STRAIN=Isolate 13;
RX MEDLINE=95012603; PubMed=7927676;
RA Sirakova T.D., Markaryan A., Kolattukudy P.E.;
RT "Molecular cloning and sequencing of the cDNA and gene for a novel
RT elastinolytic metalloproteinase from Aspergillus fumigatus and its
RT expression in Escherichia coli.";
RL Infect. Immun. 62:4208-4218(1994).
RN [2]
RP SEQUENCE OF 246-258.
RX MEDLINE=94245315; PubMed=8188335;
RA Markaryan A., Morozova I., Yu H., Kolattukudy P.E.;
RT "Purification and characterization of an elastinolytic
RT metalloprotease from Aspergillus fumigatus and immunoelectron
RT microscopic evidence of secretion of this enzyme by the fungus
RT invading the murine lung.";
RL Infect. Immun. 62:2149-2157(1994).
CC -!- FUNCTION: CATALYZES THE HYDROLYSIS OF ELASTIN.
CC -!- COFACTOR: BINDS 1 ZINC ION (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M36.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC ---
DR EMBL; L29566; AAB07708.1; -.
DR MEROPS; M36.001; -.
DR InterPro; IPR001842; Fungalysin.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF02128; Peptidase_M36; 1.
DR PRINTS; PR00999; FUNGALYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Zymogen.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 245
FT CHAIN 246 634 EXTRACELLULAR ELASTINOLYTIC
FT METALLOPROTEINASE.
FT METAL 429 429 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 430 430 BY SIMILARITY.
FT METAL 433 433 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 634 AA; 68726 MW; B94E274BF767F911 CRC64;

Query Match 3.4%; Score 103; DB 1; Length 634;
Best Local Similarity 22.8%; Pred. No. 2.7;
Matches 39; Conservative 27; Mismatches 73; Indels 32; Gaps 7;

```
Qy        312 EPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSA 371
                :|  || |       |:       | :|   :|    | |  :: |::   |       |
Db        151 DPVTALKGTTNTLQLPITV----DSASSESTEEKESYVFKGVSGTVSDPKP------KLV 200

Qy        372 YMIKDFPDFQIDVIWKYLTEVPDG--LTSAEMKDALLQVDMFGGEIHKVVWDATAVAQRE 429
                | :||  |  : : |:  |::     ||  : |          | |||  ||         :
Db        201 YFVKD--DGTLALAWRVETDIDSNWLLTYIDAKS--------GEEIHGVV---------D 241

Qy        430 YIIKLQYQTY-WQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKG 479
                |: :  || | |   |        |:| :: :   :  : | :|   : :|
Db        242 YVAEADYQVYAWGINDPTEGERTVIKDPWDSVASEFTWISDGSTNYTTSRG 292
```

RESULT 10
ELM2_ASPFU
ID ELM2_ASPFU STANDARD; PRT; 634 AA.
AC P46075;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Extracellular elastinolytic metalloproteinase precursor (EC 3.4.24.-).
GN MEP.
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=DELTA18;
RX MEDLINE=95231298; PubMed=7715453;
RA Jaton-Ogay K., Paris S., Huerre M., Quadroni M., Falchetto R.,
RA Togni G., Latge J.-P., Monod M.;
RT "Cloning and disruption of the gene encoding an extracellular
RT metalloprotease of Aspergillus fumigatus.";
RL Mol. Microbiol. 14:917-928(1994).
RN [2]
RP REVISIONS.
RC STRAIN=DELTA18;
RA Sanglard D.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- COFACTOR: BINDS 1 ZINC ION (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M36.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---

```
DR   EMBL; Z30424; CAA83015.1; -.
DR   MEROPS; M36.001; -.
DR   InterPro; IPR001842; Fungalysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF02128; Peptidase_M36; 1.
DR   PRINTS; PR00999; FUNGALYSIN.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Zymogen.
FT   SIGNAL       1     18       POTENTIAL.
FT   PROPEP      19    245
FT   CHAIN      246    634       EXTRACELLULAR ELASTINOLYTIC
FT                               METALLOPROTEINASE.
FT   METAL      429    429       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   430    430       BY SIMILARITY.
FT   METAL      433    433       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   634 AA;  68718 MW;  7BA5A9B70DE184BE CRC64;

  Query Match             3.4%;  Score 102;  DB 1;  Length 634;
  Best Local Similarity   24.0%;  Pred. No. 3.3;
  Matches   41;  Conservative   26;  Mismatches   72;  Indels   32;  Gaps    8;

Qy        312 EPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSA 371
              :|  || |        |:       | :|   :|   | |  :: |:  | |    | |
Db        151 DPVTALKGTTNTLQLPITV----DSASSESTEEKESYVFKGVSGTV--SDP----KAKLV 200

Qy        372 YMIKDFPDFQIDVIWKYLTEVPDG--LTSAEMKDALLQVDMFGGEIHKVVWDATAVAQRE 429
              | :||  |  : : |:  |::     ||  : |         | ||| ||         :
Db        201 YFVKD--DGTLALAWRVETDIDSNWLLTYIDAKS--------GEEIHGVV---------D 241

Qy        430 YIIKLQYQTY-WQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKG 479
              |: :  || | |   |        |:| :: :   :  : | :|   : :|
Db        242 YVAEADYQVYAWGINDPTEGERTVIKDPWDSVASEFTWISDGSTNYTTSRG 292


RESULT 11
AMYG_CLOSP
ID   AMYG_CLOSP     STANDARD;      PRT;   702 AA.
AC   P29761;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE   (1,4-alpha-D-glucan glucohydrolase).
GN   CGA.
OS   Clostridium sp. (strain G0005).
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1506;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92339427; PubMed=1633799;
RA   Ohnishi H., Kitamura H., Minowa T., Sakai H., Ohta T.;
RT   "Molecular cloning of a glucoamylase gene from a thermophilic
RT   Clostridium and kinetics of the cloned enzyme.";
RL   Eur. J. Biochem. 207:413-418(1992).
CC   -!- FUNCTION: CGA HAS TYPICAL KINETIC PROPERTIES FOR A GLUCOAMYLASE,
CC       BUT THIS BACTERIAL ENZYME HAD HIGHER ISOMALTOSE-HYDROLYZING
CC       ACTIVITY THAN OTHER EUKARYOTIC GLUCOAMYLASES.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC       glucose residues successively from non-reducing ends of the chains
CC       with release of beta-D-glucose.
CC   -!- SUBCELLULAR LOCATION: CELL SURFACE.
CC   -!- MISCELLANEOUS: HIGH ACTIVITY TOWARDS 1,6-GLYCOSIDIC BONDS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------

DR   EMBL; D12818; BAA02251.1; -.
DR   PIR; JQ0868; JQ0868.
DR   PIR; S24140; S24140.
DR   InterPro; IPR000165; GH_15.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR   PROSITE; PS00820; GLUCOAMYLASE; 1.
KW   Hydrolase; Glycosidase; Polysaccharide degradation; Lipoprotein;
KW   Signal.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22    702       GLUCOAMYLASE.
FT   LIPID       22     22       N-ACYL DIGLYCERIDE (PROBABLE).
FT   BINDING    342    342       SUBSTRATE (BY SIMILARITY).
FT   ACT_SITE   452    452       CATALYTIC BASE (BY SIMILARITY).
FT   ACT_SITE   455    455       GENERAL ACID CATALYST (BY SIMILARITY).
FT   ACT_SITE   456    456       INTERACT WITH SUBSTRATES (BY SIMILARITY).
SQ   SEQUENCE   702 AA;  78658 MW;  955EB4D0AD569546 CRC64;

  Query Match             3.4%;  Score 102;  DB 1;  Length 702;
  Best Local Similarity   20.4%;  Pred. No. 3.7;
  Matches  121;  Conservative   82;  Mismatches  205;  Indels  184;  Gaps   33;

Qy         21 PDKPDPRLPSMKQGFNRRWIGTNIDFVYVVY--TPQGACT-----ALDRAMEKCSPGTVR 73
              | : |    : |||     :||  ::|  |:     || :      :| |  |     ::
Db         43 PGEEDTWASAQKQG-----VGTANNYVSKVWFTLANGAISEVYYPTIDTADVK----EIK 93

Qy         74 -IVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWGSF---KTLFRDH 129
               ||: |   : || ||    |       ||   |   || : : |   | :   | :| |
Db         94 FIVTDG---KSFVSDETKDTI----SKVEKFTDKSLGYKLVNTDKK-GRYRITKEIFTD- 144

Qy        130 GRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKD 189
                                                        |    ::|   :
Db        145 ----------------------------------------------VKRNSLIMKAKFEA 158

Qy        190 SEG--NDGELFWAH----TGGGGGNFGIITKYYFKDLPMSPRG----VIASNLHFSWDGF 239
                ||  :| :|: |:       |  | | : |     :: |: |      ::||:   | |:
Db        159 LEGSIHDYKLYLAYDPHIKNQGSYNEGYVIKANNNEMLMAKRDNVYTALSSNI--GWKGY 216

Qy        240 T------RDALQDL-----LTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSND- 287
              :      | : ||     :||::  ||    | :   :|  :   :| : |    | |
Db        217 SIGYYKVNDIMTDLDENKQMTKHYDSAR---GNIIEGAEIDLKKNSQFEIVLSFGNSEDE 273

Qy        288 ---AEREVAQDRHYHLE-ADIEQIYKTCEPTKALGGHAGWAPFP----VRPRKRHTSKTS 339
                 |  |   : :  |: | |::  | |       | |    :     ::  :  |:| :
Db        274 AVKASIETLSENYDSLKSAYIDEWEKYCNSLNNFNGKANSLYYNSMMILKASEDKTNKGA 333

Qy        340 YMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIW-KYLTEVPDGLTS 398
              |:   ::  |:       :| | :  | |               ::| : |  | :   :
Db        334 YI--ASLSIPWG------DGQGDDNTGGY-------------HLVWSRDLYHVANAFIA 371

Qy        399 A-EMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQ-EEDKDAVNLKW---- 452
              | ::  |  :|     : ||| |   : |  :|    | | | :| | : | :      :
Db        372 AGDVDSANRSLDY----LAKVVKDNGMIPQNTWISGKPYWTGIQLDEQADPIILSYRLRR 427

Qy        453 ----------IRDFY--------EEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVD-- 492
                       : ||        :| :|  ||   | |      |:    |    : |
Db        428 YDLYDSLVKPLADFIIKMGPKTGQERWEEIGGY-SPATMAAEVAGLTCAAYIAEQNKDYE 486

Qy        493 --------LNNWKN-------GKYGALELYFLGNLNRLIKAKWLWDPNEIFT 529
                      :||:        ::| ||      |    |:   | |||  ||
Db        487 SAQKYQEKADNWQKLIDNLTYTEHGPLE-----NGQYYIRIAGLPDPNADFT 533


RESULT 12
DPOM_NEUIN
ID   DPOM_NEUIN     STANDARD;      PRT;   969 AA.
AC   P33538;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Probable DNA polymerase (EC 2.7.7.7).
OS   Neurospora intermedia.
OG   Mitochondrion.
OG   Plasmid kalilo.
```

```
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5142;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92035090; PubMed=1934129;
RA   Chan B.S.S., Court D.A., Vierula P.J., Bertrand H.;
RT   "The kalilo linear senescence-inducing plasmid of Neurospora is an
RT   invertron and encodes DNA and RNA polymerases.";
RL   Curr. Genet. 20:225-237(1991).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG, TO
CC       DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52106; CAA36327.1; ALT_SEQ.
DR   PIR; S17909; S17909.
DR   InterPro; IPR002064; DNA_pol_B.
DR   InterPro; IPR004868; DNA_pol_B_2.
DR   Pfam; PF03175; DNA_pol_B_2; 1.
DR   PRINTS; PR00106; DNAPOLB.
DR   SMART; SM00486; POLBc; 1.
DR   PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   DNA-binding; Plasmid; Mitochondrion.
SQ   SEQUENCE   969 AA;  112829 MW;  6048B993814F668F CRC64;

  Query Match             3.4%;  Score 101.5;  DB 1;  Length 969;
  Best Local Similarity   20.0%;  Pred. No. 6.3;
  Matches   76;  Conservative   50;  Mismatches  151;  Indels  103;  Gaps   16;

Qy        213 TKYYFKDLPMSPRGVIASNLHFSW---------DGFTRDALQDLLTKYFKLARCDWKNTV 263
              :|:|      || | | |    ||          |  ||::  | |  :  | :  :|
Db         38 SKPYSTSTD-SPMGHIVSASKSSWYNYYSHKNYDPLTRESFHDELRGFISLYKKQFKE-- 94

Qy        264 GKFQIFHQAAEEFVMYLYTSYSNDAER-------------------EVAQDRHYHLEADI 304
                  |:|        ::|:  |                     |     : |    |
Db         95 ----------EKFFFMAKIKFNNNDIRSISTVQIGSTDPLEVLRLLEAITSTYMHTHTGI 144

Qy        305 EQIYKTCEPTK----ALGG--HAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETIN 358
               :     || :    :||     |   | :|    : || | |   :        : ||
Db        145 SE--AVSEPFEYELFSLGDGLPKGNIIFTFKPTSNPSIKTKYEHKSNI-----KRNKNIN 197

Qy        359 GSGPNQRGKYK-SAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQ-VD----MFG 412
               |  |   |:| : | | :  |       |  :  : ||  :  : : :   ||     |
Db        198 LSKKNPLNKFKYNGYTIPNTMDLS---QWPNIHFINDGKNAVSLNNIIKSGVDNMTLSFF 254

Qy        413 GEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNT 472
               |:|   : | :     | |:: :    |:|  :       :  :::|
Db        255 ITINKKYNEITVLLNNTPIFKIKDEKIMSEDDLSSFKRTITENEQDKVY---------- 303

Qy        473 QVESGKGVF--EGCYFNY----PDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNE 526
               |:|: ||  |    ::       || |::| |   |:|
Db        304 VFENGEMVFFSENVKTSFIKKITRQDLINFENPKIITLDL-------------------- 343

Qy        527 IFTNKQSIPTKPLKEPKQTK 546
                 :|:|  |:|| |  |
Db        344 ---ETRSVPIHPIKEGKDGK 360


RESULT 13
MCEL_SFVKA
ID   MCEL_SFVKA     STANDARD;      PRT;  836 AA.
AC   P25950;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   mRNA capping enzyme large subunit [Includes: Polynucleotide 5'-
DE   triphosphatase (EC 3.1.3.33) (mRNA 5'-triphosphatase) (TPase); mRNA
DE   guanylyltransferase (EC 2.7.7.50) (GTP--RNA guanylyltransferase)
DE   (GTase)].
GN   D3R.
OS   Shope fibroma virus (strain Kasza) (SFV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Leporipoxvirus.
OX   NCBI_TaxID=10272;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91306463; PubMed=1649507;
RA   Upton C., Stuart D., McFadden G.;
RT   "Identification and DNA sequence of the large subunit of the capping
RT   enzyme from Shope fibroma virus.";
RL   Virology 183:773-777(1991).
CC   -!- FUNCTION: CATALYZES THE FIRST TWO REACTIONS IN THE MRNA CAP
CC       FORMATION PATHWAY.
CC   -!- CATALYTIC ACTIVITY: 5'-phosphopolynucleotide + H(2)O =
CC       polynucleotide + phosphate.
CC   -!- CATALYTIC ACTIVITY: GTP + (5')PP-pur-mRNA = diphosphate +
CC       G(5')PPP-pur-mRNA.
CC   -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.
CC   -!- SIMILARITY: BELONGS TO THE VIRAL GTASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63902; AAA47224.1; -.
DR   PIR; A40478; QQVZRA.
DR   InterPro; IPR004971; Pox_MCEL.
DR   Pfam; PF03291; Pox_MCEL; 1.
KW   Transferase; Nucleotidyltransferase; Hydrolase; mRNA processing;
KW   mRNA capping; Multifunctional enzyme.
FT   ACT_SITE    256    256       GUANYLYLATION SITE (POTENTIAL).
SQ   SEQUENCE   836 AA;  97018 MW;  7375FDA548AE1730 CRC64;

  Query Match             3.3%;  Score 97.5;  DB 1;  Length 836;
  Best Local Similarity   18.3%;  Pred. No. 11;
  Matches   86;  Conservative   75;  Mismatches  169;  Indels  141;  Gaps   21;

Qy         28 LPSMKQGFNRRWIGTNIDFVYVVYTPQGACTALDRAMEKCSPGTVRIVSGGHCYEDFVFD 87
              |||  :|            | : |: |          ::  :   : |:       |  :|
Db        365 LPSQPEG-----------VVLFYSDQRNQPDYKIKLDNTTDHMINIIYRYMSSEPVIFG 412

Qy         88 ECVKAIINVTGLVESGYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYS------- 140
              |      | | |    : ||:|:       ::|: | :  |: | |    | :
Db        413 E------NSTFLEYKKFSDDKGF-----PKDYGTGKLMLTDNVRYLNNIYCIAFTNVYED 461

Qy        141 VGLGGHIVGGGDGILARLHGLPVDWLSGVEV---VVKPVLTEDSVLKYVHKDSEGNDGEL 197
              ||:  :|          |: ::|       ::|| :  |   ||::|:  ||  ::
Db        462 VGIKNVVV----------PIKFISEFSATGELIKPRI--DKTFKYLYKEYYGNQYQI 506

Qy        198 FWAHTGGGGGNFG-IITKYYFKDLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLAR 256
              ||          | :: :    |:           |:: | :  :       |
Db        507 VVAHIRDQNIKIGDVLDEDKLSDV----------GQHYANDKYRLNPDVSYFTN------ 550

Qy        257 CDWKNTVGKFQIFHQAAEEFVMYLY---TSYSNDAEREV-------AQDRHYHLEADIEQ 306
                 | | |   |      :  :: ||   |   |  :|:|           |   :   :|
Db        551 ---KRTRGPLGILSNYVKTLLISLYCSKTFLDNSNKRKVLAIDFGNGADLEKYFYGEISS 607

Qy        307 IYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETING----SGP 362
              :  |    :|:|           |  :|: |  | :  ::  | |  : |||      |
Db        608 LVATDPDKEAIG----------RCIERYNSLNSGI--KSKYYKFDYIQETIRSVTYVSSV 655
```

```
Qy        363 NQ---RGKYKSAYMIKDFPDFQIDVIW----KYLTEVPDGLTSAEMKDALLQVDMFGGE- 414
              :    ||:       |  |:|  : :    |:   | : ||        : :    |:
Db        656 REVFFFGKF-------DLVDWQFAIHYSFHPKHYATVMNNLTELTASGGKVLITTMDGDL 708

Qy        415 -----------IHK------------------VVWDATAVAQ--REYIIK 433
                        |||                    :|:: ::::: :|||:|
Db        709 LSQLTDKKTFVIHKNLPSSENYMSVEKIHEDQILVYNPSSMSRPMQEYIVK 759


RESULT 14
MM01_BOVIN
ID   MM01_BOVIN     STANDARD;      PRT;   469 AA.
AC   P28053;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Interstitial collagenase precursor (EC 3.4.24.7) (Matrix
DE   metalloproteinase-1) (MMP-1) (Fibroblast collagenase).
GN   MMP1 OR CLG.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Periodontium fibroblast;
RX   MEDLINE=95201294; PubMed=7894061;
RA   Tamura M., Shimokawa H., Sasaki S.;
RT   "Primary structure of bovine interstitial collagenase deduced from
RT   cDNA sequence.";
RL   DNA Seq. 5:63-66(1994).
CC   -!- FUNCTION: CLEAVES COLLAGENS OF TYPES I, II, AND III AT ONE SITE IN
CC       THE HELICAL DOMAIN. ALSO CLEAVES COLLAGENS OF TYPES VII AND X.
CC   -!- CATALYTIC ACTIVITY: Cleaves preferentially one bond in native
CC       collagen. Cleavage of the triple helix of collagen at about three-
CC       quarters of the length of the molecule from the N-terminus, at
CC       775-Gly-|-Ile-776 in the alpha-1(I) chain. Cleaves synthetic
CC       substrates and alpha-macroglobulins at bonds where P1' is a
CC       hydrophobic residue.
CC   -!- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.
CC   -!- ENZYME REGULATION: CAN BE ACTIVATED WITHOUT REMOVAL OF THE
CC       ACTIVATION PEPTIDE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.
CC   -!- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58256; CAA41210.1; -.
DR   PIR; S14654; KCBOI.
DR   HSSP; P03956; 1HFC.
DR   MEROPS; M10.001; -.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR001818; Matrixin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF00045; hemopexin; 4.
DR   Pfam; PF00413; Peptidase_M10; 1.
DR   PRINTS; PR00138; MATRIXIN.
DR   SMART; SM00120; HX; 4.
DR   SMART; SM00235; ZnMc; 1.
DR   PROSITE; PS00024; HEMOPEXIN; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00546; CYSTEINE_SWITCH; 1.
KW   Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;
KW   Collagen degradation; Extracellular matrix; Signal.
FT   SIGNAL        1     18       BY SIMILARITY.
FT   PROPEP       19     99       ACTIVATION PEPTIDE.
FT   CHAIN       100    469       INTERSTITIAL COLLAGENASE.
FT   DOMAIN      275    469       HEMOPEXIN-LIKE.
FT   SITE         92     92       CYSTEINE SWITCH (POTENTIAL).
FT   METAL       218    218       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    219    219       BY SIMILARITY.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       228    228       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    278    466       PROBABLE.
FT   CARBOHYD    120    120       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   469 AA;  53354 MW;  B4A5504CE24BD7B5 CRC64;

  Query Match             3.2%;  Score 97;  DB 1;  Length 469;
  Best Local Similarity   19.4%;  Pred. No. 5.3;
  Matches   61;  Conservative   62;  Mismatches   98;  Indels   94;  Gaps   18;

Qy        180 DSVLKYV---HKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSW 236
              | :: :|   |:|:   ||          |||     : |:     |   | : ||
Db        158 DIMISFVRGDHRDNSPFDGP---------GGNLA----HAFQ-----PGAGIGGDAHFDD 199

Qy        237 DGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDR 296
              | :     ||     :::|  :: :::|   : |  : :    :| ||:   : :::||
Db        200 DEWWTSNFQDY--NLYRVAAHEFGHSLG---LAH--STDIGALMYPSYTFSGDVQLSQD- 251

Qy        297 HYHLEAD-IEQIYKTCE-PTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALT 354
                  : | |: ||   : ||: :|        |  |    :  |           | |:|
Db        252 ----DIDGIQAIYGPSQNPTQPVG--------PQTPEVCDSKLT----------FDAIT 288

Qy        355 ETINGSGPNQRGKYKSAYMIKD--FPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFG 412
              || |     : ::   ||  :  :|: ::: |  :  ::|:|| :
Db        289 -TIRGEVMFFKDRF---YMRTNPLYPEVELNFISVFWPQLPNGLQA------------- 330

Qy        413 GEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYG------G 466
                    | || |: :   :   ||  : :|      :| :  ::|  :|
Db        331 ---------AYEVADRDEVRFFKGNKYWAVKGQDV-----LRGYPRDIYRSFGFPRTVKS 376

Qy        467 VPDPNTQVESGKGVF 481
                  :    :: ::||  |
Db        377 IDAAVSEEDTGKTYF 391


RESULT 15
AGAB_VIBS7
ID   AGAB_VIBS7     STANDARD;      PRT;   955 AA.
AC   P48840;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Beta-agarase B (EC 3.2.1.81).
GN   AGAB.
OS   Vibrio sp. (strain JT0107).
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=47913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94250684; PubMed=8193156;
RA   Sugano Y., Matsumoto T., Noma M.;
RT   "Sequence analysis of the agaB gene encoding a new beta-agarase from
RT   Vibrio sp. strain JT0107.";
RL   Biochim. Biophys. Acta 1218:105-108(1994).
CC   -!- FUNCTION: HYDROLYZES AGAROSE TO YIELD PREDOMINANTLY
CC       NEOAGAROTETRAOSE AND NEOAGAROHEXAOSE.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,3-beta-D-galactosidic linkages
CC       in agarose, giving the tetramer as the predominant product.
CC   -!- SIMILARITY: BELONGS TO FAMILY 50 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; D21202; BAA04744.1; -.
KW   Hydrolase; Glycosidase; Multigene family.
SQ   SEQUENCE   955 AA;  105976 MW;  6E54555DDA8BD215 CRC64;

  Query Match             3.2%;  Score 97;  DB 1;  Length 955;
  Best Local Similarity   20.5%;  Pred. No. 14;
  Matches   56;  Conservative   46;  Mismatches   97;  Indels   74;  Gaps   16;

Qy        104 YDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPV 163
              ||: :  :|::|   |     :| || |:  |   :                  :|  |
Db        629 YDNKKVAYVANG---W-----IFGDHARISTGNDYW----------------GPIHD-PF 663

Qy        164 D--WLSGVEVVVKPVLTEDSVLKYVHKDSEGN-----DGELFWAHTGGGGGNFGIITKYY 216
              |  ::: |: : | ::||      | |:          | |: | :|     ::|::
Db        664 DPEFVNSVKAMTKKLMTE------VDKNDPWMMGVFVDNEISWGNTKNDANHYGLVVNAL 717

Qy        217 FKDLPMSP-RGVIASNLHFS-WDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAE 274
              |:  || :     :|    |      |::|| |        |   |  :  | :: :
Db        718 SYDMKKSPAKAAFTEHLKEKYW------AIEDLNT--------SWGVKVASWAEFEKSFD 763

Qy        275 EFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKT--CEPTKALGGH----AGWAPFPV 328
                    : :   : : :|    ||    : : |   |  | | |   |    | :| : |
Db        764 ---------HRSRLSKNMKKDYAEMLEMLSAKYFSTVRAELKKVLPNHLYLGAPFADWGV 814

Qy        329 RPRKRHTSKTSYMHDETMDYPFYALTETINGSG 361
              |     :| :  : : | |  ||  | :|  |
Db        815 TP---EIAKGAAPYVDVMSYNLYA--EDLNSKG 842


Search completed: June 10, 2003, 09:59:16
Job time : 27 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 09:56:56 ; Search time 22 Seconds
(without alignments)
2385.882 Million cell updates/sec

Title: US-09-998-284-2
Perfect score: 2997
Sequence: 1 MATLPQKDPGYIVIDVNAGT.....IFTNKQSIPTKPLKEPKQTK 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	736.5	24.6	685	2	AG0038	probable oxidoredu
2	368.5	12.3	530	2	F86390	hypothetical prote
3	360.5	12.0	367	2	F69816	reticuline oxidase
4	347.5	11.6	539	2	T10626	reticuline oxidase
5	346	11.5	528	2	T10622	hypothetical prote
6	333.5	11.1	529	2	A86391	hypothetical prote
7	328.5	11.0	535	2	E86390	hypothetical prote
8	324.5	10.8	541	2	G86151	F22M8.11 protein -
9	322	10.7	540	2	T10625	reticuline oxidase
10	302.5	10.1	552	2	H86390	T1K7.21 protein -
11	300.5	10.0	532	2	T00461	probable berberine
12	300	10.0	532	2	T10624	reticuline oxidase
13	298.5	10.0	527	2	G86390	T1K7.22 protein -
14	296	9.9	536	2	F86251	hypothetical prote
15	290.5	9.7	535	2	T07969	probable reticulin
16	286.5	9.6	538	2	A41533	reticuline oxidase
17	278.5	9.3	447	2	G70034	reticuline oxidase
18	278	9.3	533	2	A86433	T5I8.19 protein -
19	277.5	9.3	526	2	H86432	hypothetical prote
20	277	9.2	527	2	E86432	T5I8.15 protein -
21	272.5	9.1	530	2	T10628	hypothetical prote
22	252	8.4	531	2	F86432	T5I8.16 protein -
23	249.5	8.3	540	2	T00463	probable berberine
24	242.5	8.1	466	2	T49756	related to berberi
25	240	8.0	527	2	G86432	hypothetical prote
26	238.5	8.0	461	2	A70687	probable oxidoredu
27	207.5	6.9	479	2	H70847	probable oxidoredu
28	203.5	6.8	479	2	B95997	probable oxidoredu
29	190.5	6.4	805	2	AI0176	probable oxidoredu
30	185	6.2	431	2	D86433	hypothetical prote
31	180.5	6.0	446	2	G70510	probable oxidoredu
32	154.5	5.2	448	2	A55519	mcrA protein - Str
33	154	5.1	123	2	G70969	hypothetical prote
34	140.5	4.7	461	2	H69350	glycolate oxidase
35	139.5	4.7	511	2	T48777	6-HYDROXY-D-NICOTI
36	118.5	4.0	769	2	B87681	tyrosine kinase Di
37	116	3.9	478	1	KCRBS1	stromelysin 1 (EC
38	113	3.8	592	2	E70455	sulfur oxidation p
39	111	3.7	1009	2	C64483	hypothetical prote
40	110	3.7	781	2	T36143	probable secreted
41	109.5	3.7	496	2	S50518	hypothetical prote
42	109	3.6	934	2	T42394	potassium channel
43	108	3.6	1599	2	S22737	glucosyltransferas
44	108	3.6	2817	2	B97033	uncharacterized pr
45	106.5	3.6	1050	2	JC7578	endo-1,4-beta-xyla

ALIGNMENTS

RESULT 1
AG0038
probable oxidoreductase YPO0308 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AG0038
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.,
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; H
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AG0038
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-685 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89170.1; PID:g15978409; GSPDB:GN00175
C;Genetics:
A;Gene: YPO0308

Query Match 24.6%; Score 736.5; DB 2; Length 685;
Best Local Similarity 33.6%; Pred. No. 3.2e-48;
Matches 195; Conservative 77; Mismatches 205; Indels 103; Gaps 20;

```
Qy         13 VIDVNAGTPDKPDPRLPSMKQGFNRRW---IGTNIDFVYVVYTPQGACTALDRAMEKCSP 69
              :|| |  | :       ::::||| ||   :    : :|:  ||      | : |:   :
Db          3 IIDKNVSTYE-------TLQKGFNLRWPPNVEQGAETIYICTTPDEVFAATNTALAAGNR 55

Qy         70 GTVRIVSGGHCYEDFVFDECVKAIINVTGLVE-SGYDDDRGYFVSS-------------- 114
               |||  ||||||| || ::     :::  | | || | |    ::|
Db         56 ITVR--SGGHCYEGFVSNKLSTERLSIIDLGEMSGLDYDEDKTITSLWDANKNTYRFKSL 113

Qy        115 -GDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVV 173
               |: ||  : :|::  || :||||||||:|||| ||| |:|:|||| |||::||:::|
Db        114 TGNQNWNGYVSLYKRSGRTIPGGSCYSVGVGGHISGGGYGLLSRLHGLTVDWVTGVDILV 173

Qy        174 KPVLTEDSVLKYVHKDSEGN-DGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNL 232
                    ::|  ||    | ||  |  | ||||||||  ||| ||| :|:      |
Db        174 PVGNAHRLAFRHVRADSVSEVDRELLMACCGAGGGNFGIIIAYYFDDLPKAPQKAYWIPL 233

Qy        233 HFSWDGFTRDALQDLLTKYFKLARCDW----------KNTVGKFQIF------H-QAAE 274
               : |       :|:     : | |   |          |  ||   :|      |  |::
Db        234 TYPW-----SSLKATFPAFLK-AYWQWFADNDVNATSTKEGVGNGGLFTLLKLNHIDASD 287

Qy        275 EFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAP-------FP 327
               |: :  :  |    :|       |   ||:          :   ||  |       |
Db        288 NVVLAIQYTGPNG---QVGGANDIPLNDFIEK----------MNAAAGMTPTIYDDFILP 334

Qy        328 VRPRKRHT---SKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDV 384
               |  :|     |     ||:    :  :|: ||||| |||||||| | || | |   ::
Db        335 NIPPFKHLYPGRKIGRTVDESASMDWLHVTQMINGSGSNQRGKYKSDYQIKQFSD---EM 391
```

```
Qy        385 IWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEED 444
              || :            :|:|:| :|| |:     ||||:||  ::| ||||||  |
Db        392 CHALLTHLTTATADKRFNQSLVQIDSYGGAINSRGIGATAVSQRNSLLKAQYQTYWTNEA 451

Qy        445 KDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLN---------N 495
               |  :| |||: |  :   : | | |           |||||  ||||:|:         |
Db        452 DDQTHLTWIRNIYAAV---HNGKPAPPE--------FEGCYINYPDIDMKYTDSGEEDPN 500

Qy        496 WKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIP 535
              | |  ||    :    : |||  |   ||| || :: |||
Db        501 WLNLYYG----WDTQLIKRLIALKARIDPNNIFHHELSIP 536


RESULT 2
F86390
hypothetical protein T1K7.23 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: F86390
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F86390
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-530 <STO>
A;Cross-references: GB:AE005172; NID:g9797759; PIDN:AAF98577.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: poppy reticuline oxidase

  Query Match             12.3%;  Score 368.5;  DB 2;  Length 530;
  Best Local Similarity   27.0%;  Pred. No. 3.1e-20;
  Matches  134;  Conservative   76;  Mismatches  180;  Indels  107;  Gaps   23;

Qy         72 VRIVSGGHCYEDFVFDECVKAII-------NVTGLVESGYDDDRGYFVSSGDTNWGSFKT 124
              :|| |||| ||    :    |  :|        ::|  | |      :  :: :| |   |
Db        103 LRIRSGGHDYEGLSYMSSVPFVILDMYNLRSITVDVSS-----KKAWIQAGAT-LGELYT 156

Qy        125 LFRDHGRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSV 182
               |  : |   | | | :|| |||| ||| | | | :|: || :   :::       |
Db        157 NVNDVSQTLAFPAGVCATVGAGGHISGGGYGNLMRKYGITVDHVIDAQII-------DVN 209

Qy        183 LKYVHKDSEGNDGELFWAHTGGGGGNFGIIT--KYYFKDLPMSPRGVIASNLHFSWDGFT 240
               | ::: : | |  ||||  |||||:||:|    |     |:|     |    |      |
Db        210 GKLLNRATMGED--LFWAIRGGGGGSFGVILSWKINLVDVPKIVT-VFKVNKTLEQGG-- 264

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEE----------FVMYLYTSYSND 287
              |:| |        |:    ||  :| :|  :                :  |  :
Db        265 ----TDVLYK--------WQLVASKFPESLFVRAMPQVANGTKRGERTITVVFYAQFLGR 312

Qy        288 AEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMD 347
               :  :|       |  ::  :: |:     |      || :|        : || :    :|
Db        313 TDALMAIMNQNWPELGLK--HEDCQEMSWLNSTLFWADYPA------GTPTSIL----LD 360

Qy        348 YPFYALTETINGSGPNQRGKYKSAYMIKDPPDFQIDVIWKYLTEVPDGLTSAEMKDALLQ 407
              |          | |    | || |: |  |   :: :|| : :  : :          :|
Db        361 RP----------SSPGDFFKSKSDYVKKPIPKEGLEKLWKTMLKFNNNI-------VWMQ 403

Qy        408 VDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGG 466
               : :|| : ::   |||   |:  : |:|| | |   :    :|  ::: | |: |||
Db        404 FNPYGGVMDRIPATATAFPHRKGNLPKIQYFTTWFNANATMSSLSQMKELY-EVAEPY-- 460

Qy        467 VPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGK--------YGALELYFLGNLNRLIKA 518
                    | |         :||| |:|: :  :|:         ||:   |||||| ||:
Db        461 -------VSSNP---REAFFNYRDIDVGSNPSGETNVDEAKIYGS--KYFLGNLKRLMDV 508

Qy        519 KWLWDPNEIFTNKQSIP 535
                  |  :||:  | |:||||
Db        509 KAKYDPDNFFKNEQSIP 525


RESULT 3
F69816
reticuline oxidase homolog ygaK - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: F69816
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berterc
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F69816
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <KUN>
A;Cross-references: GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CAB12708.1; PID:e1182869;
A;Experimental source: strain 168
C;Genetics:
A;Gene: ygaK

  Query Match             12.0%;  Score 360.5;  DB 2;  Length 367;
  Best Local Similarity   26.0%;  Pred. No. 7.5e-20;
  Matches  118;  Conservative   61;  Mismatches  157;  Indels  117;  Gaps   15;

Qy        106 DDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDW 165
              | :  :: :|       ::||:: :|  || |:  :|| |  :||| |:| |  ||  |
Db          9 DKKLAYIEAGAELGEVYRTLWQ-YGLTLPAGTIANVGLTGLTLGGGIGLLTRAAGLTCDS 67

Qy        166 LSGVEVVVKPVLTEDSVLKYVHKDSEG---------NDGELFWAHTGGGGGNFGIITKYY 216
              |  :|::|                : ||          |  :||||  ||||||||||:|
Db         68 LVQLEMIVA-------------DEKEGADLITVSCSNHPDLFWASQGGGGGNFGIVTSMT 114

Qy        217 FKDLPMSPRGVIASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEF 276
              || :|:|    :   :: : || |                    |:|
Db        115 FKAVPISQVSIF--SITWGWDDFEE-------------VYNTWQN---------------- 144

Qy        277 VMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFP-------VR 329
                 :  |::|             | : ||    |      :|||   |    |       ::
Db        145 ----WPPYTDD-----------RLTSSIEFWPKEVNRIEALGQFVG----PKTELKKLLK 185

Qy        330 PRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYK--SAYMIKDFPDFQIDVIWK 387
              |  :  | || |   |   ||        |  | ||  | |   ::: |   :  | |
Db        186 PLLKAGSPTSGMVKTT---PFIEAVTFFNSPGGNQPQKMKRSGSFIEKPLSERAISTIKH 242

Qy        388 YLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDA 447
              :|   |:       ::| :     ||   :|  | ||   |: || :| | |    :
Db        243 FLEHAPN-------QNASVWQQALGGAAGRVAPDQTAFYYRDAIIAQEYLTNWTSPGEKR 295

Qy        448 VNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGALELY 507
              |::||      : :                 | | |:||::: ||    ||
Db        296 QNVRWIEGLRTSLSKE----------------TMGDYVNWPDIEIRNWPRTYYGE---- 334

Qy        508 FLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540
                |: || : |  :||  :|  :|||| ||:
Db        335 ---NVERLRRVKTTYDPENVFRFEQSIP--PLR 362
```

RESULT 4
T10626
reticuline oxidase homolog F21C20.190 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Jun-2001
C;Accession: T10626
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16991
A;Accession: T10626
A;Molecule type: DNA
A;Residues: 1-539 <BEV>
A;Cross-references: EMBL:AL080254; GSPDB:GN00062; ATSP:F21C20.190
A;Experimental source: cultivar Columbia; BAC clone F21C20
C;Genetics:
A;Gene: ATSP:F21C20.190
A;Map position: 4
C;Superfamily: poppy reticuline oxidase

Query Match 11.6%; Score 347.5; DB 2; Length 539;
Best Local Similarity 26.2%; Pred. No. 1.3e-18;
Matches 141; Conservative 76; Mismatches 181; Indels 141; Gaps 28;

```
Qy          52 TPQGACTALDR------AMEKCSPG---TVRIVSGGHCYE--DFVFDECVKAIINVTGLV 100
               ||: |     |      |   ||      ::| |||| ||   :: |:    |:::: |
Db          83 TPKPAIIVTPRSDIHVSAAVTCSKSLNFLLKIRSGGHDYEGLSYISDKPF-FILDMSNLR 141

Qy         101 ESGYD-DDRGYFVSSGDT-------NWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGD 152
               :   |  |:  ::|:| |         |    |      ||   | | | :||:|||| |||
Db         142 DVSVDIADQSAWISAGATLGEVYYRIWEKSKV----HG--FPAGVCPTVGVGGHISGGGY 195

Qy         153 GILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGII 212
               | : |  || || |   ::|       |   : : : | | |  |||| :|||| :||::
Db         196 GNMLRKFGLSVDNLIDAKIV-------DVNGQILDRKSMGED--LFWAISGGGGASFGVV 246

Qy         213 TKYYFKDLPMSPRGVI----------ASNLHFSWDG---------FTRDALQDLLTKYFK 253
               |  | :|:     :          | ::   |           | |  :| :  | |
Db         247 LGYKVKLVPVPETVTVFRVEKYMDSGAVDMVHKWQSVGPKTDRNLFLRMLIQPVTRKKVK 306

Qy         254 LARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEP 313
               |     ||    :|   ||| |  |   :               |  :::  : |
Db         307 TVRA----TV--VALFLGRAEEVVALLGKEFP---------------ELSLKK--ENCSE 343

Qy         314 TKALGGHAGW----APFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYK 369
                     |     |  : |:           |  :|   :                 || |
Db         344 MTWFQSALWWDNRVNPTQIDPK--------VFLDRNLDRANF--------------GKRK 381

Qy         370 SAYMIKDFPDFQIDVIWKYLTEVPD-GLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQR 428
               | |:  : |   |: ::| :||:   ||           : :||:: :|  :||    |
Db         382 SDYVASEIPRDGIESLFKKMTELGKIGLV----------FNPYGGKMAEVTVNATPFPHR 431

Qy         429 EYIIKLQYQTYWQEEDKDAVNLKWIRDFYEE---MYEPYGGVPDPNTQVESGKGVFEGCY 485
               : |:||   |||   ::| ::  : |  :   :|    |    | :             |
Db         432 SKLFKIQYSVTWQE---NSVEIE--KGFLNQANVLYSFMTGFVSKNPR---------NAY 477

Qy         486 FNYPDVDL-------NNWKNGK-YGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPT 536
               || |||:        |::: |: ||    ||  | :||:| |   ||:  | |:|||||
Db         478 LNYRDVDIGVNDHGTNSYEEGEVYG--RKYFGDNFDRLVKVKTAADPDNFFRNEQSIPT 534
```

RESULT 5
T10622
hypothetical protein F21C20.150 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Jun-2001
C;Accession: T10622
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16991
A;Accession: T10622
A;Molecule type: DNA
A;Residues: 1-528 <BEV>
A;Cross-references: EMBL:AL080254; GSPDB:GN00062; ATSP:F21C20.150
A;Experimental source: cultivar Columbia; BAC clone F21C20
C;Genetics:
A;Gene: ATSP:F21C20.150
A;Map position: 4
C;Superfamily: poppy reticuline oxidase

Query Match 11.5%; Score 346; DB 2; Length 528;
Best Local Similarity 27.0%; Pred. No. 1.6e-18;
Matches 134; Conservative 68; Mismatches 187; Indels 108; Gaps 23;

```
Qy          72 VRIVSGGHCYEDFVFDECVK-AIINVTGLVESGYDDDRGY-FVSSGDTNWGSFKTLFRDH 129
               :|| |||| ||   |   |   |::: |     |  |   :| :| |  |   |
Db         102 IRIRSGGHDYEGLSFTSSVPFVILDMHDLRSITIDVFRKQAWVDAGAT-MGELYTKIAAA 160

Qy         130 GRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVH 187
                : |   || | ::| |||| ||| | | | :|: || :    :|
Db         161 SKTLAFAGGVCPTLGAGGHISGGGYGNLIRKYGISVDHVVDARIV--------------- 205

Qy         188 KDSEGN-------DGELFWAHTGGGGGNFGIIT--KYYFKDLPMS----------PRGVI 228
               |  ||          :| ||  |||| :||:|   |    |:| :          :||
Db         206 -DVNGNILTGATLGRDLLWAIRGGGGASFGVILSWKINLVDVPKTVTVFKVNKTLEQGV- 263

Qy         229 ASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDA 288
               ::: : |   :    |||  :          |   |       : :   :  |  :   |
Db         264 -TDVLYKWQLVSSKLPQDLFLRAMP------KPVNGVV----PSEKTIAVVFYAQFLGSA 312

Qy         289 EREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDY 348
               | :|       |  :::  : |    :     |  :||       : || :    :|
Db         313 RRLMAIMNKNLPELGLKR--EDCYEMSWINTTTFWQNYPV------GTSTSVL----LDR 360

Qy         349 PFYALTETINGSGP-NQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQ 407
               |          |||      | || |: |  |  ::: |||  :          :  :  :|
Db         361 P----------SGPAGAFYKSKSDYVKKPIPKEEMEKIWKAM---------LKFNNMWMQ 401

Qy         408 VDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGG 466
                : :|| : |:  ||||    |:  : |:||   | : :    ||  :|| | || |||
Db         402 WNPYGGVMDKIPADATAFPHRKGNLFKIQYFALWTDANATYANLGLMRDIYHEM-EPY-- 458

Qy         467 VPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGK--------YGALELYFLGNLNRLIKA 518
                       | |         : || |:|: :  :|:         ||:   |||||  ||::
Db         459 -------VSSNP---REAFLNYRDIDVGSNPSGETNLEEAKIYGS--KYFLGNFKRLMEV 506

Qy         519 KWLWDPNEIFTNKQSIP 535
               |  :||    |  :||||
Db         507 KAKYDPENFFRFEQSIP 523
```

RESULT 6
A86391
hypothetical protein T1K7.20 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: A86391
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A86391
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-529 <STO>
A;Cross-references: GB:AE005172; NID:g9797756; PIDN:AAF98574.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

```
C;Superfamily: poppy reticuline oxidase

  Query Match           11.1%;  Score 333.5;  DB 2;  Length 529;
  Best Local Similarity  25.8%;  Pred. No. 1.5e-17;
  Matches  128;  Conservative   77;  Mismatches  184;  Indels  107;  Gaps   22;

Qy         72 VRIVSGGHCYEDFVFDECVKAII-------NVTGLVESGYDDDRGYFVSSGDTNWGSFKT 124
              :|| ||||  |   :   |  :|       ::|  | |     :  :| :| |  |*
Db        103 LRIRSGGHDLEGLSYRSSVPFVILDMFNLRSITVNVLS-----KKAWVQAGAT-LGELYV 156

Qy        125 LFRDHGRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSV 182
                 :  : |  | | | :||:|||| ||| | | |  |: || :|  :::       |
Db        157 KINEASQTLAFPAGVCPTVGVGGHISGGGYGNLMRKFGITVDHVSDAQLI-------DVN 209

Qy        183 LKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPR--GVIASNLHFSWDGFT 240
              | ::: | | |  ||||  |||| :||:|  :   :|   |:   |    |       |
Db        210 GKLLNRASMGED--LFWAIRGGGGASFGVILSWKI-NLVKVPKILTVFKVNKTLEQGG-- 264

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEEFV----------MYLYTSYSND 287
                |:| |         |:    ||   :| :|  : :           :  |  :
Db        265 ----TDVLYK--------WQLVATKFPEDLFMRAWPQIINGAERGDRTIAVVFYAQFLGP 312

Qy        288 AEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMD 347
              |:: :|       |  : :  : |          || :|    |      | : |   :
Db        313 ADKLLAIMNQRLPELGLRR--EDCHEMSWFNTTLFWADYPAGTPK------SVLLDRPTN 364

Qy        348 YPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDAL-L 406
              |:          | || |: |  |   :: :||          |  :  : : :
Db        365 PGFF---------------KSKSDYVKKPIPKEGLEKLWK---------TMFKFNNIVWM 400

Qy        407 QVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYG 465
              | : :|| : ::   |||   |:  : |:|| | |   :   ::|  ::: | :: |||
Db        401 QFNPYGGVMDQIPSTATAFPHRKGNMFKVQYSTTWLAANATEISLSMMKELY-KVAEPY- 458

Qy        466 GVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGALEL------YFLGNLNRLIKAK 519
                | |        :||| |:|: :  : :    |       ||||| ||:: |
Db        459 --------VSSNP---REAFFNYRDIDIGSNPSDETNVDEAKIYGYKYFLGNLKRLMQVK 507

Qy        520 WLWDPNEIFTNKQSIP 535
                :||   | |:||||
Db        508 AKYDPENFFKNEQSIP 523


RESULT 7
E86390
hypothetical protein T1K7.24 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: E86390
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E86390
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-535 <STO>
A;Cross-references: GB:AE005172; NID:g9797760; PIDN:AAF98578.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: poppy reticuline oxidase

  Query Match           11.0%;  Score 328.5;  DB 2;  Length 535;
  Best Local Similarity  27.5%;  Pred. No. 3.6e-17;
  Matches  138;  Conservative   66;  Mismatches  180;  Indels  117;  Gaps   25;

Qy         72 VRIVSGGHCYEDFVFDECVKAII-------NVTGLVESGYDDDRGYFVSSGDTNWGSFKT 124
              :|| ||||  |   :   |  :|       ::|  | |     :  :| :| |  |
Db        103 IRIRSGGHDNEGLSYVSSVPFVILDMHKLRDITVDVSS-----KKAWVQAGAT-LGELYV 156

Qy        125 LFRDHGRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSV 182
                 :  : |  | | | :|| |||| ||| | | |  | || :  |:|       |
Db        157 KIDEASQTLAFPAGICATVGAGGHISGGGYGNLMRKFGTTVDHVIDAELV-------DVN 209

Qy        183 LKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPR--GVIASNLHFSWDGFT 240
              | ::: : | |  ||||  |||| :||:|  :   :|   |:   |    |       |
Db        210 GKLLNRSTMGED--LFWAIRGGGGASFGVILSWKI-NLVEVPKIFTVFQVNKTLEQGG-- 264

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEEFVMYLYTSYSNDAEREVA----- 293
                |:: |         |:    ||   :| :|  : |      : :   || :|
Db        265 ----TDVVYK--------WQLVANKFPDNLFLRAMPQVV-----NGTKHGERTIAIVFWA 307

Qy        294 --QDRHYHLEADIEQIY-------KTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDE 344
                |    |   : | :       : |:     |      ||  |       | ||  :
Db        308 QFLGRTDELMEIMNQSFPELGLRREDCQEMSWLNTTLFWAMLPA-----GTPKTVLLGRP 362

Qy        345 TMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDA 404
              | |  |:                 | || |: |  |   :: |||           |  :  :
Db        363 T-DPVFF--------------KSKSDYVKKPIPKEGLEKIWK---------TMLKFNNI 397

Qy        405 L-LQVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYE 462
              : |  : :|| : ::  :|||   |:  : |:|| | | : :    ||  ::: | |: |
Db        398 VWLHFNPYGGMMDRIPSNATAFPHRKGNLFKVQYYTTWLDPNATESNLSIMKELY-EVAE 456

Qy        463 PYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGK--------YGALELYFLGNLNR 514
              ||          | |        :||| |:|: :  :|:          ||    ||||| |
Db        457 PY---------VSSNP---REAFFNYRDIDIGSNPSGETDVDEAKIYG--YKYFLGNLKR 502

Qy        515 LIKAKWLWDPNEIFTNKQSIP 535
              |:  |   ||    | |:||||
Db        503 LMDVKAKSDPENFFKNEQSIP 523


RESULT 8
G86151
F22M8.11 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C;Accession: G86151
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86151
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-541 <STO>
A;Cross-references: GB:AE005172; NID:g8570449; PIDN:AAF76476.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: poppy reticuline oxidase

  Query Match           10.8%;  Score 324.5;  DB 2;  Length 541;
  Best Local Similarity  25.4%;  Pred. No. 7.3e-17;
  Matches  132;  Conservative   73;  Mismatches  169;  Indels  145;  Gaps   26;

Qy         76 SGGHCYED---------FVFDECVKAIINVTGLVESGYDDDRGYFVSSGDT-------NW 119
              |||| |:          || |  :  : |:|  |:   ||    :| :| |       |
Db        110 SGGHDYDGVSYISNRPFFVLD--MSYLRNIT--VDMS-DDGGSAWVGAGATLGEVYYNIW 164

Qy        120 GSFKTLFRDHG-RVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLT 178
                | ||    ||    | | | :|| |||| ||| | : | :|| ||:::  ::|
```

```
Db        165 QSSKT----HGTHGFPAGVCPTVGAGGHISGGGYGNMIRKYGLSVDYVTDAKIV------ 214

Qy        179 EDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVI---------- 228

Db        215 -DVNGRILDRKSMGED--LFWAIGGGGGASFGVILSFKIKLVPVPPRVTVFRVEKTLVEN 271

Qy        229 ASNLHFSW---------DGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMY 279

Db        272 ALDMVHKWQFVAPKTSPDLFMRLMLQPVT----------RNTT-------QTVRASVVA 313

Qy        280 LYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWA----PFPVRPR---K 332

Db        314 LFLGKQSDLMSLLTKE---FPELGLKP--ENCTEMTWIQSVMWWANNDNATVIKPEILLD 368

Qy        333 RHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEV 392

Db        369 RNPDSASFL------------------------KRKSDYVEKEISKDGLDFLCKKLMEA 403

Qy        393 PD-GLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVN-- 449

Db        404 GKLGLV----------FNPYGGKMSEVATTATPFPHRKRLFKVQHSMNWKDPGTDVESSF 453

Qy        450 LKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDL-------NNWKNGK-Y 501

Db        454 MEKTRSFYSYM-APFVTKNPRHT-----------YLNYRDLDIGINSHGPNSYREAEVY 500

Qy        502 GALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540

Db        501 G--RKYFGENFDRLVKVKTAVDPENFFRDEQSIPTLPTK 537
```

RESULT 9
T10625
reticuline oxidase homolog F21C20.180 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Jun-2001
C;Accession: T10625
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16991
A;Accession: T10625
A;Molecule type: DNA
A;Residues: 1-540 <BEV>
A;Cross-references: EMBL:AL080254; GSPDB:GN00062; ATSP:F21C20.180
A;Experimental source: cultivar Columbia; BAC clone F21C20
C;Genetics:
A;Gene: ATSP:F21C20.180
A;Map position: 4
C;Superfamily: poppy reticuline oxidase

Query Match 10.7%; Score 322; DB 2; Length 540;
Best Local Similarity 24.4%; Pred. No. 1.1e-16;
Matches 125; Conservative 79; Mismatches 172; Indels 136; Gaps 22;

```
Qy         72 VRIVSGGHCYE--DFVFDECVKAIINVTGLVESGYD-DDRGYFVSSGDTNWGSFKTLF-- 126

Db        113 LKIRSGGHDYDGLSYISDKPF-FILDMSNIRDVSVDIASNSAWISAGATLGEVYYRIWEK 171

Qy        127 -RDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKY 185

Db        172 SRVHG--FPAGVCPTVGVGGHLSGGGYGNMVRKFGLSVDYVEDAKIV-------DVNGRV 222

Qy        186 VHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVI----------ASNLHFS 235

Db        223 LDRKAMGED--LFWAITGGGGGSYGVVLGYKVKLVPVPSVVTVFRVEQYMDSGAVDMVHK 280

Qy        236 WDG---------FTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSN 286

Db        281 WQSVGPKTDPNLFMRMLIQPVTRKKVKTVR------ASVVALFLGRADEVVALLSKEF-- 332

Qy        287 DAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMH---- 342

Db        333 ------------------------PELGL-------------KKENCSEMTWFQSALW 353
```

```
Qy        343 -----DETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPD-GL 396

Db        354 WDNRLNATQVDPKVFLDRNLDTS---SFGKRKSDYVATAIPKKGIESLFKKMIELGKIGL 410

Qy        397 TSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEE----DKDAVNLKW 452

Db        411 V----------FNPYGGKMAEVAVNAKPFPHRNKLFKIQYSVNWKENSAEIEKGYLNQAK 460

Qy        453 IRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDL-------NNWKNGK-YGAL 504

Db        461 V------LYSFMTGFVSKNPR---------SSYFNYRDVDIGVNDHGANSYKEGEVYG-- 503

Qy        505 ELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPT 536

Db        504 RKYFGENFDRLVKIKTAVDPGNFFRNEQSIPT 535
```

RESULT 10
H86390
T1K7.21 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C;Accession: H86390
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86390
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-552 <STO>
A;Cross-references: GB:AE005172; NID:g9797757; PIDN:AAF98575.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: poppy reticuline oxidase

Query Match 10.1%; Score 302.5; DB 2; Length 552;
Best Local Similarity 25.3%; Pred. No. 3.6e-15;
Matches 128; Conservative 72; Mismatches 190; Indels 115; Gaps 24;

```
Qy         72 VRIVSGGHCYEDFVFDECVK-AIINVTGLVESGYDDDR-GYFVSSGDTNWGSFKTLFRDH 129

Db        127 LRIRSGGHDNEGFSYMSSVPFVILDMHNLRSIDVNLSRKNAWVQAGAT-LGELYVKINEA 185

Qy        130 GRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVH 187

Db        186 SQTLAFPAGVCPTVGAGGHISGGGFGNLMRKFGITVDHVIDAQII-------DVNGKLLN 238

Qy        188 KDSEGNDGELFWAHTGGGGGNFGIITKY-----------YFKDLPMSPRGVIASNLHFS 235

Db        239 RAAMGED--LFWA-IRGGGSSFGVILSWKINLVEVPKILTVFKVNKTLEQG--GTDILYK 293

Qy        236 WDGFTRDALQDLLTKYFKLARCDWKNTV-----GKFQIFHQAAEEFVMYLYTSYSNDAER 290

Db        294 WQ-LVANKLPDSL---FITA---WPRTVNGPKPGERTV------AVVFYAQFLGPTDKLM 340

Qy        291 EVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKR------HTSKTSYMHDE 344

Db        341 EIMDQSFPELGLGRED----CHEMSWLNTTLFWANYPAGTPKSILLDRPPTNSVSF---- 392

Qy        345 TMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDA 404

Db        393 ----------------------KSKSDFVKKPIPKKGLEKLWKTMFKFNSSVS------- 423

Qy        405 LLQVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEP 463
```

```
Db        424 -LQFNPYGGVMDRIPATATAFPHRKGNLFKVQYSTMWFDANATESSLAMMNELF-EVAEP 481

Qy        464 YGGVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGK--------YGALELYFLGNLNRL 515

Db        482 Y---------VSSNP---REAFFNFRDIDIGSNPSGETNVDEAKIYGS--KYFLGNLKRL 527

Qy        516 IKAKWLWDPNEIFTNKQSIPTKPLK 540

Db        528 MDVKAKYDPDNFFKNEQSIPPVRVK 552
```

RESULT 11
T00461
probable berberine bridge enzyme [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F19I3.2
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 15-Jun-2001
C;Accession: T00461; H84760
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, April 1998
A;Description: Arabidopsis thaliana chromosome II BAC F19I3 genomic sequence.
A;Reference number: Z14160
A;Accession: T00461
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-532 <ROU>
A;Cross-references: EMBL:AC004238; NID:g3033373; PID:g3033375
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: H84760
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-532 <STO>
A;Cross-references: GB:AE002093; NID:g3033375; PIDN:AAC12819.1; GSPDB:GN00139
C;Genetics:
A;Gene: F19I3.2; At2g34790
A;Map position: 2
A;Introns: 164/2; 326/3
C;Superfamily: poppy reticuline oxidase

  Query Match 10.0%; Score 300.5; DB 2; Length 532;
  Best Local Similarity 24.0%; Pred. No. 4.9e-15;
  Matches 117; Conservative 88; Mismatches 197; Indels 85; Gaps 22;

```
Qy         72 VRIVSGGHCYEDFVF---DECVKAIINVTGLVESGYD-DDRGYFVSSGDT---NWGSFKT 124

Db        108 LRLRSGGHDYEGLSFVAEDETPFVIVDLSKLRQVDVDLLDSNSAWAHAGATIGEVYYRIQE 167

Qy        125 LFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLK 184

Db        168 KSQTHG--FPAGLCSSLGIGGHLVGGAYGSMMRKFGLGADNVLDARIV-------DANGQ 218

Qy        185 YVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFTRDAL 244

Db        219 ILDRAAMGED--VFWAIRGGGGGSFGVILAWKIKLVPV-PATVTVFTV-------TKTLE 268

Qy        245 QDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAE---REVAQDRHYHLE 301

Db        269 QDGTKVLYK-----WEQIADKLD-----DDLFIRVIISPASKTTKPGNRTISMSYQAQFL 318

Qy        302 ADIEQIYKTCEPTKALGGHAGWAPFP-VRPRKRHTSKTSYMHD--ETMDYPFYALTETI- 357

Db        319 GDSNRLLQVMQKS-----------FPELGLTKKDCTEMSWIKSVMYIAGFPNSAAPEALL 367

Qy        358 -NGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQV-DMFGGEI 415

Db        368 AGKSLFKNHFKAKSDFVKEPIPVEGLEGLWERFLE----------EDSPLTIWNPYGGMM 417

Qy        416 HKVVWDATAVAQRE-YIIKLQYQTYWQE-EDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQ 473

Db        418 SRISESEIPFPHRNGTLFKIQWLSTWQDGKVSEERHMKWIREMYSYMEQYVSKNP----- 472

Qy        474 VESGKGVFEGCYFNYPDVDLNNWKNGKYGALE---LYFLGNLNRLIKAKWLWDPNEIFTN 530

Db        473 --------RQAYVNYRDLDLGT-NEGETDAREWGAKYYKGNFERLVKIKGEFDPDNFFRH 523

Qy        531 KQSIPTK 537

Db        524 EQSVPTK 530
```

RESULT 12
T10624
reticuline oxidase homolog F21C20.170 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Jun-2001
C;Accession: T10624
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke,
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16991
A;Accession: T10624
A;Molecule type: DNA
A;Residues: 1-532 <BEV>
A;Cross-references: EMBL:AL080254; GSPDB:GN00062; ATSP:F21C20.170
A;Experimental source: cultivar Columbia; BAC clone F21C20
C;Genetics:
A;Gene: ATSP:F21C20.170
A;Map position: 4
C;Superfamily: poppy reticuline oxidase

  Query Match 10.0%; Score 300; DB 2; Length 532;
  Best Local Similarity 23.5%; Pred. No. 5.3e-15;
  Matches 129; Conservative 78; Mismatches 196; Indels 146; Gaps 24;

```
Qy         36 NRRWIGTNIDFVYVVYTPQGACTALDRAMEKCSPG---TVRIVSGGHCYE--------DF 84

Db         73 NQRFSAPDVPKPVLILTPVQPSDV--QSAVKCARRFGIHIRTRSGGHDYEGLSYVTHKPF 130

Qy         85 VFDECVKAIINVTGLVESGYD-DDRGYFVSSGDTNWGSFKTLFRDHGR-----VLPGGSC 138

Db        131 V-------ILDLRNLRSITVDVDNRSVWVQTGAT----IGELYYEIGKKNRTLAFPAGVC 179

Qy        139 YSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELF 198

Db        180 PTVGVGGHFSGGGYGTLLRKHGLAADHVIDARVV-------DARGRILERREMGED--FF 230

Qy        199 WAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFS----------WDGFTRDALQD-- 246

Db        231 WAIRGGGGSSFCVVLSWKIGLINVPSTVTVFNVTKFSEQSALKIIHRWQ-FVADKVSDDL 289

Qy        247 ----LLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEA 302

Db        290 FIRVMLQRYKNMVRASFPG---------------LYL-GSVKNLLKMVNKEFPELGLEE 332

Qy        303 D-------IEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTE 355

Db        333 DDCTEMSWIESVIWFAE----LG---------------------------EEPINVLTK 360

Qy        356 TINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEI 415

Db        361 RTRAS---LAFKAKSDFVQEPMPKTAISKLWRRLQE-------PEAEHAQLIFTPFGGKM 410

Qy        416 HKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQV 474

Db        411 SEIADYETPFPHRKGNIYEIQYLNYWRGDVKEKY-MRWVERVYDDMSEFVAKSP------ 463

Qy        475 ESGKGVFEGCYFNYPDVDL--------NNWKNGKYGALELYFLGNLNRLIKAKWLWDPNE 526

Db        464 -------RGAYINLRDLDLGMYVGVKRSKYEEGKSWGVK-YFKNNFERLVRVKTSVDPSD 515
```

```
Qy        527 IFTNKQSIP 535
                | ::||||
Db        516 FFCDEQSIP 524


RESULT 13
G86390
T1K7.22 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C;Accession: G86390
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86390
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-527 <STO>
A;Cross-references: GB:AE005172; NID:g9797758; PIDN:AAF98576.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: poppy reticuline oxidase

  Query Match             10.0%;  Score 298.5;  DB 2;  Length 527;
  Best Local Similarity   25.4%;  Pred. No. 6.9e-15;
  Matches  138;  Conservative   68;  Mismatches  200;  Indels  137;  Gaps   27;

Qy         48 YVVYTPQGACTALD----------------RAMEKCSPGT---VRIVSGGHCYEDFVFDE 88
              || ||     ::|:                :|    |:      :|| |||| ||   :
Db         60 YVSYTKNKRYSSLNYQKLVAIVAAKHVSHVQATVVCAKANGIQLRIRSGGHDYEGLSYTS 119

Qy         89 CVKAII-------NVTGLVESGYDDDRGYFVSSGDTNWGSFKTLFRDHGRVL--PGGSCY 139
               |  :|       ::|  | |      :  :| :| |  |   |   :  : |  | | |
Db        120 SVPFVILDMYNLRSITVDVSS-----KKAWVQAGAT-LGELYTKINEASQTLAFPAGVCP 173

Qy        140 SVGLGGHIVGGGDGILARLHGLPVDWLSGVEVV-VKPVLTEDSVLKYVHKDSEGNDGELF 198
              :||:|||| ||| | | |  |: || :   ::: |   | : : :        | |  ||
Db        174 TVGVGGHITGGGFGNLMRKFGITVDHVIDAQLIGVNGKLLDRATM--------GED--LF 223

Qy        199 WAHTGGGGGNFGIITKY------------YFKDLPMSPRGVIASNLHFSWDGFTRDALQD 246
              ||  |||| :||:|  :             ||      :|   ::: : |         :|
Db        224 WAIRGGGGASFGVILSWKINLVEVPKILTVFKVSKTLEQG--GTDVLYKWQLVATKVPED 281

Qy        247 L-------LTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYH 299
              |       : |  ||     : |:|   :|       |  :|  :    |   |:
Db        282 LFIRAWPQIVKGTKLG----ERTIG--VVF------FAQFLGPT---DKLMEIMSQSLPE 326

Qy        300 LEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETING 359
              |     |      |              || :||   |                :| |
Db        327 LGLRRED----CHEMSWFNTTLFWANYPVGTPTR----------VLLDRP---------- 362

Qy        360 SGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVV 419
              | | :  |/ ||  : |  |   :: |||          |  :     :: : :|| : ::
Db        363 STPGEFFKSKSDNIKKPIPKEGLEKIWK---------TMLKFNFVWIEFNPYGGVMDRIP 413

Qy        420 WDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGK 478
              |||   |:  :  ||| | | :  :    |  ::: | |:  ||          | |
Db        414 ATATAFPHRKGNLFNLQYSTIWLDAKETENKLTMMKELY-EVAGPY---------VSSNP 463

Qy        479 GVPEGCYFNYPDVDLN------NWKNGK-YGALELYFLGNLNRLIKAKWLWDPNEIFTNK 531
                  ||: | |:       |    | ||    |||||| ||:  |   ||:  | |:
Db        464 ---REALFNFRDFDIGINPSGLNVDEAKIYG--YKYFLGNLKRLMDVKAKCDPDNFFKNE 518

Qy        532 QSI 534
              |||
Db        519 QSI 521


RESULT 14
F86251
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 15-Jun-2001
C;Accession: F86251
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F86251
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-536 <STO>
A;Cross-references: GB:AE005172; NID:g4835780; PIDN:AAD30246.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: poppy reticuline oxidase

  Query Match             9.9%;  Score 296;  DB 2;  Length 536;
  Best Local Similarity   24.2%;  Pred. No. 1.1e-14;
  Matches  125;  Conservative   65;  Mismatches  190;  Indels  136;  Gaps   20;

Qy         76 SGGHCYED---------FVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWGSFKTLF 126
              |||| ||           || |     | |     || :       | :| |    :  ::
Db        110 SGGHDYEGVSYISHVPFFVLDMSNLRNITVDPATESAW-------VGAGATLGEVYYRIW 162

Qy        127 ---RDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVL 183
                 : ||   | | | :|| |||| ||| | : | :|| ||:::  ::|          ||
Db        163 EKTKSHG--FPAGVCPTVGAGGHISGGGYGNMIRKYGLSVDYVTDAKIVD----VNGQVL 216

Qy        184 KYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVI----------ASNLH 233
                | :|   ::|||  |||| :||:|  :  | :|: |   :           |: :
Db        217 -----DRKGMGEDMFWAINGGGGASFGVILAFKIKLVPVPPTVTVFRVEKNLVENATEMV 271

Qy        234 FSWDG---------FTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSY 284
               |            | |  || :     :  |                       |: |:
Db        272 HKWQFVAPKTDPGLFMRLLLQPVTRNKMQTVRAS-----------------VVALFLGD 313

Qy        285 SNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRH--TSKT---S 339
               |     : ::                          ||    |:   |  |   |
Db        314 QNTVMSMLTKE----------------------------FPELGLKKENCTEMTWIQS 343

Qy        340 YMHDETMDYPFYALTETINGSGPNQR--GKYKSAYMIKDFPDFQIDVIWKYLTEVPD-GL 396
               |      |       | :    |:    || || :: |:     :| ::| : ||   ||
Db        344 VMWWANNDNATQIKPEILLDRNPDMATFGKRKSDFVEKEITKDGLDFLFKKMIEVGKIGL 403

Qy        397 TSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVN--LKWIR 454
                     : :|| :  |    |    |: : |:|:   |::   :|    |: :
Db        404 V----------FNPYGGIMSTVATTKTPFPHRKKLYKIQHSMNWKDPGTEAETSFLQKAK 453

Qy        455 DFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDL--NNWKNGKYGALE----LYF 508
               ||  |  |:      :|              | || |:|:  |      |   |    :||
Db        454 SFYSYM-APFVTKNPRHT-----------YINYRDLDIGVNTPGPNSYRVAEVFGRMYF 500

Qy        509 LGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLKEPKQ 544
               | :||:| |   ||   | ::|||| | |  ::
Db        501 GENFDRLVKVKTAVDPQNFFRDEQSIPTLPGKPARR 536


RESULT 15
T07969
probable reticuline oxidase (EC 1.5.3.9) - opium poppy
```

```
N;Alternate names: berberine bridge enzyme
C;Species: Papaver somniferum (opium poppy)
C;Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 15-Jun-2001
C;Accession: T07969
R;Facchini, P.J.; Penzes, C.; Johnson, A.G.; Bull, D.
Plant Physiol. 112, 1669-1677, 1996
A;Title: Molecular characterization of berberine bridge enzyme genes from opium poppy.
A;Reference number: Z16255; MUID:97127853; PMID:8972604
A;Accession: T07969
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-535 <FAC>
A;Cross-references: EMBL:AF025430; NID:g3282516; PIDN:AAC61839.1; PID:g3282517
C;Genetics:
A;Gene: bbe1
C;Superfamily: poppy reticuline oxidase
C;Keywords: oxidoreductase

  Query Match             9.7%;  Score 290.5;  DB 2;  Length 535;
  Best Local Similarity   23.8%;  Pred. No. 2.9e-14;
  Matches  123;  Conservative   87;  Mismatches  160;  Indels  147;  Gaps   28;

Qy         71 TVRIVSGGHCYE--------DFVFDECVKA------IINVTGLVESGYDDDRGYFVSSGD 116
              |:|: |||| ||         ||  : :         ::: |  ||||      |:  :
Db        100 TIRLRSGGHSYEGLSYTADTPFVIVDMMNLNRISIDVLSETAWVESGATLGELYYAIAQS 159

Qy        117 TNWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPV 176
              |:   |             | | :|| |||| ||| |:::| :||  |       ||  :
Db        160 TDTLGF-----------TAGWCPTVGSGGHISGGGFGMMSRKYGLAAD------NVVDAI 202

Qy        177 LTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVI-------- 228
              |  ||    : ::  |:|  :|||  ||||| :| |  :  | ||:  :  :
Db        203 LI-DSNGAILDREKMGDD--VFWAIRGGGGGVWGAIYAWKIKLLPVPEKLTVFRVTKNVG 259

Qy        229 ---ASNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYS 285
                ||:|   |  :  | | :  |          : :|         : ::|:|
Db        260 IEDASSLLHKWQ-YVADELDEDFT----------VSVLGGV----NGNDAWLMFLGLHLG 304

Qy        286 -NDAEREVAQDRHYHL---EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYM 341
               || : :  ::   |   : : :::              |
Db        305 RKDAAKTIIDEKFPELGLVDKEFQEM-------------SWG----------------- 333

Qy        342 HDETMDYPFYALTETINGSGPNQRGKY-KSAYMIK-DFPDFQIDV-IWKYLTEVPDGLTS 398
              |:|   | :  :||:    |:  |: : |:  | ||    : : ::::  |:
Db        334 --ESM--AFLSGLDTISELN-NRFLKFDERAFKTKVDFTKVSVPLNVFRHALEM-----L 383

Qy        399 AEMKDALLQVDMFGGEIHKVVWDATAVAQR-------EYIIKLQYQTYWQEEDKDAVN-- 449
              :|     : :: |||:: ::  | |    |       ||||       | ::::  :
Db        384 SEQPGGFIALNGFGGKMSEISTDFTPFPHRKGTKLMFEYIIA------WNQDEESKIGEF 437

Qy        450 LKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLN--NWKN--GKYGALE 505
              :|:  ||: :         :|    |    |     | |: |:|:   :|:|      |:|
Db        438 SEWLAKFYDYL--------EPFVSKEPRVG-----YVNHIDLDIGGIDWRNKSSTTNAVE 484

Qy        506 L-------YFLGNLNRLIKAKWLWDPNEIFTNKQSIP 535
              :       ||  |  ||:||| | ||| :| : ||||
Db        485 IARNWGERYFSSNYERLVKAKTLIDPNNVFNHPQSIP 521

Search completed: June 10, 2003, 10:00:28
Job time : 25 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 09:49:11 ; Search time 43 Seconds
(without alignments)
1691.973 Million cell updates/sec

Title: US-09-998-284-2
Perfect score: 2997
Sequence: 1 MATLPQKDPGYIVIDVNAGT.....IFTNKQSIPTKPLKEPKQTK 546

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2997	100.0	546	22	AAU02192	Synthetic hexose o
2	2997	100.0	546	22	AAB59205	Chondrus cripus he
3	2989	99.7	546	18	AAW20076	Hexose oxidase, an
4	2985	99.6	546	21	AAY83619	Synthetic hexose o
5	635.5	21.2	539	20	AAY39306	SpnJ protein invol
6	635.5	21.2	539	22	AAB70952	S. spinosa protein
7	368.5	12.3	530	23	ABB91135	Herbicidally activ
8	347.5	11.6	539	23	ABB93022	Herbicidally activ
9	333.5	11.1	529	23	ABB91138	Herbicidally activ
10	328.5	11.0	535	23	ABB91134	Herbicidally activ
11	324.5	10.8	541	23	ABB90805	Herbicidally activ
12	322	10.7	540	23	ABB93021	Herbicidally activ
13	306.5	10.2	545	21	AAY88391	Tetrahydrocannabin
14	302.5	10.1	552	23	ABB91137	Herbicidally activ
15	300	10.0	532	23	ABB93020	Herbicidally activ
16	298.5	10.0	527	23	ABB91136	Herbicidally activ
17	296	9.9	536	23	ABB90968	Herbicidally activ
18	294.5	9.8	458	21	AAG41845	Arabidopsis thalia
19	294.5	9.8	541	21	AAG41844	Arabidopsis thalia
20	294.5	9.8	556	21	AAG41843	Arabidopsis thalia
21	292.5	9.8	509	19	AAW55062	Arabidopsis antifu
22	290.5	9.7	535	22	AAB61789	Opium poppy reticu
23	286.5	9.6	538	22	AAB61790	California poppy r
24	278	9.3	533	23	ABB91198	Herbicidally activ
25	277.5	9.3	526	23	ABB91197	Herbicidally activ
26	277	9.2	508	19	AAW55061	Arabidopsis antifu
27	277	9.2	527	23	ABB91194	Herbicidally activ
28	277	9.2	529	22	AAB61787	Sunflower carbohyd
29	275	9.2	529	19	AAW55053	Sunflower leaf ant
30	274	9.1	529	22	AAB61788	Sunflower carbohyd
31	272.5	9.1	525	21	AAG43378	Arabidopsis thalia
32	272.5	9.1	530	21	AAG43377	Arabidopsis thalia
33	272	9.1	529	19	AAW55054	Sunflower leaf ant
34	266	8.9	540	19	AAW55059	Lettuce antifungal
35	265.5	8.9	544	22	AAB81081	Cannabidiolic acid
36	264.5	8.8	421	21	AAG43379	Arabidopsis thalia
37	263	8.8	508	19	AAW55060	Arabidopsis antifu
38	255	8.5	495	20	AAY24919	Microdochium nival
39	252	8.4	531	23	ABB91195	Herbicidally activ
40	249.5	8.3	474	21	AAG29445	Arabidopsis thalia
41	249.5	8.3	540	21	AAG29444	Arabidopsis thalia
42	249.5	8.3	551	21	AAG29443	Arabidopsis thalia
43	240	8.0	527	21	AAG42979	Arabidopsis thalia
44	240	8.0	527	23	ABB91196	Herbicidally activ
45	240	8.0	534	21	AAG42978	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAU02192
ID AAU02192 standard; Protein; 546 AA.
XX
AC AAU02192;
XX
DT 12-SEP-2001 (first entry)
XX
DE Synthetic hexose oxidase (HOX) amino acid sequence.
XX
KW HOX; seaweed; hexose oxidase; interleukin I receptor antagonist; IL-1ra;
KW food manufacturing; beverage; detergent; baking; dough improving agent;
KW D-hexose:O2-oxireductase; EC 1.1.3.5.
XX
OS Chondrus crispus.
OS Synthetic.
XX
PN WO200138544-A1.
XX
PD 31-MAY-2001.
XX
PF 24-NOV-2000; 2000WO-IB01886.
XX
PR 24-NOV-1999; 99GB-0027801.
XX
PA (DANI-) DANISCO AS.
XX
PI Johansen CL, Kjaerulff S, Madrid SM, Pedersen H, Poulsen CH;
PI Zargahi MR;
XX
DR WPI; 2001-367695/38.
DR N-PSDB; AAS06173.
XX

PT Releasing soluble or membrane associated intracellular protein from a
PT cell for manufacturing food, comprises contacting the cell with a
PT membrane extracting composition and causing the protein to be released
PT in soluble form -
XX
PS Claim 16; Fig 6; 108pp; English.
XX
CC The sequence represents the amino acid sequence of synthetic hexose
CC oxidase (D-hexose:O2-oxireductase, EC 1.1.3.5), also referred to as HOX.
CC The native HOX gene was altered using site-directed mutagenesis in order
CC to match the codon usage to known codon preferences of biotechnologically
CC relevant yeasts, such as Pichia sp., to facilitate high level production
CC of HOX in these organisms. The invention involves a method for releasing
CC a soluble or membrane associated intracellular protein of interest (POI)
CC from a cell involving contacting a cell comprising a soluble or membrane
CC associated intracellular POI with a membrane extracting composition (I)
CC and causing the POI to be released from the cell in a soluble form. The
CC method is useful for releasing POI, such as an interleukin I receptor
CC antagonist (IL-1ra) which involves contacting a transformed cell
CC comprising IL-1ra with (I) and causing IL-1ra to be released from the
CC transformed cell, in a soluble form. The method is also useful for
CC screening mutated cells or transformed cells producing elevated levels of
CC intracellular POI. The method is used to release a POI for manufacturing
CC food products, such as beverages, preparation of detergents, and in
CC baking as a dough improving agent. The method obtains a fast, specific
CC and economically efficient extraction of a soluble or membrane associated
CC intracellular POI without the use of conventional cell disruption
CC techniques. The resulting cell extract contains less contaminating
CC intracellular DNA and is relatively free of cell wall fragments. The
CC intracellular POI can be recovered from a eukaryotic host organism such
CC as yeast, before glycosylation takes place. The method can be used to
CC prevent contact of intracellular POI with the extracellular growth
CC medium.
XX
SQ Sequence 546 AA;

Query Match 100.0%; Score 2997; DB 22; Length 546;
Best Local Similarity 100.0%; Pred. No. 3.9e-281;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60

Qy         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300

Qy        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360

Qy        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420

Qy        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480

Qy        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540

Qy        541 EPKQTK 546
              ||||||
Db        541 EPKQTK 546
```

RESULT 2
AAB59205
ID AAB59205 standard; Protein; 546 AA.
XX
AC AAB59205;
XX
DT 23-MAR-2001 (first entry)
XX
DE Chondrus cripus hexose oxidase enzyme protein.
XX
KW Hexose oxidase; marine alga; anti-fouling.
XX
OS Chondrus cripus.
XX
PN WO200075293-A2.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-IB00829.
XX
PR 04-JUN-1999; 99GB-0013050.
XX
PA (DANI-) DANISCO AS.
XX
PI Poulsen CH, Kragh KM;
XX
DR WPI; 2001-112148/12.
XX
PT New anti-fouling composition, useful as a coating for treating
PT different surfaces, e.g. outdoor woodwork, external surface of a
PT central heating system, or a hull of a marine vessel -
XX
PS Claim 5; Page 35-36; 36pp; English.
XX
CC The present invention relates to a new anti-fouling composition. The
CC composition involves a surface coating material, a hexose oxidase
CC enzyme obtained from a marine organism and a substrate for the
CC enzyme. The anti-fouling composition is useful as a coating formulated
CC for treating a surface, e.g. outdoor wood work, external surface of a
CC central heating system, or a hull of a marine vessel. It is
CC also useful as an anti-fouling agent for marine structures exposed to
CC seawater flora and fauna.
XX
SQ Sequence 546 AA;

Query Match 100.0%; Score 2997; DB 22; Length 546;
Best Local Similarity 100.0%; Pred. No. 3.9e-281;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60

Qy         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240
```

```
Qy        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300

Qy        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360

Qy        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420

Qy        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480

Qy        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540

Qy        541 EPKQTK 546
              ||||||
Db        541 EPKQTK 546
```

RESULT 3
AAW20076
ID AAW20076 standard; Protein; 546 AA.
XX
AC AAW20076;
XX
DT 12-SEP-1997 (first entry)
XX
DE Hexose oxidase, an antimicrobial/antioxidant agent.
XX
KW Hexose oxidase; Chondrus crispus; marine algae; recombinant;
KW antimicrobial; antioxidant; food preparation; dairy product; starch;
KW beverage; animal feed; silage; sugar reduction; cosmetics; dental;
KW toothpaste; dough; lactone production.
XX
OS Chondrus crispus.
XX
PN WO9640935-A1.
XX
PD 19-DEC-1996.
XX
PF 04-JUN-1996; 96WO-DK00238.
XX
PR 07-JUN-1995; 95US-0476910.
XX
PA (BIOT-) BIOTEKNOLOGISK INST.
XX
PI Hansen OC, Stougaard P;
XX
DR WPI; 1997-052332/05.
DR N-PSDB; AAT76552.
XX
PT Recombinant production of polypeptide having hexose oxidase activity
PT - used in food preparations as antibacterial and antioxidant agent
XX
PS Example 3.5; Page 95-97; 124pp; English.
XX
CC AAW20076 shows the hexose oxidase (HO) enzyme of the marine algae
CC species Chondrus crispus. HO is useful in the production of food
CC products, e.g. dairy products, starch-containing food products (dough)
CC and non-dairy beverages. HO and active peptide fragments have
CC antimicrobial and antioxidant properties and act by removing all the
CC oxygen in a food packaging. HO and peptides of HO can also be used in an
CC animal feed, especially silage. Further uses are to reduce, or analyse,
CC the sugar content in a food, in the production of cosmetics, tooth care
CC products or a pharmaceutical product and in lactone production. HO can
CC be recombinantly produced in industrially appropriate quantities, and at
CC a quality and purity level which renders the polypeptide suitable for
CC industrial purposes.
XX
SQ Sequence 546 AA;

```
  Query Match             99.7%;  Score 2989;  DB 18;  Length 546;
  Best Local Similarity   99.8%;  Pred. No. 2.4e-280;
  Matches  545;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db          1 MATLPQKDPGYIVIDVNAGTADKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60

Qy         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300

Qy        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360

Qy        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420

Qy        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480

Qy        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540

Qy        541 EPKQTK 546
              ||||||
Db        541 EPKQTK 546
```

RESULT 4
AAY83619
ID AAY83619 standard; Protein; 546 AA.
XX
AC AAY83619;
XX
DT 29-AUG-2000 (first entry)
XX
DE Synthetic hexose oxidase.
XX
KW Hexose oxidase; production; fermentation; synthetic; modification;
KW prokaryote; eukaryote.
XX
OS Synthetic.
XX
PN EP1008651-A2.
XX
PD 14-JUN-2000.
XX
PF 01-DEC-1999; 99EP-0204068.
XX
PR 09-DEC-1998; 98DK-0001630.
XX
PA (BIOT-) BIOTEKNOLOGISK INST.

XX
PI Stougaard P, Pedersen LH, Wolff AM, Poulsen U, Hansen OC;
XX
DR WPI; 2000-389309/34.
DR N-PSDB; AAZ94011.
XX
PT Nucleic acid fragment useful for producing large amounts of hexose
PT oxidase comprises nucleotide sequence encoding hexose oxidase that is
PT modified by at least one codon
XX
PS Disclosure; Page 28-29; 42pp; English.
XX
CC Wild type hexose oxidase can be produced by isolating a nucleotide
CC sequence coding for hexose oxidase naturally produced by an organism
CC and modifying the sequence so that the modified sequence is
CC expressed, under identical conditions, at a level at least 10% higher
CC than the non-modified sequence. The method is useful for producing
CC hexose oxidase in prokaryotic and eukaryotic cells. Previous methods
CC of producing hexose oxidase were to isolate the enzyme from a source
CC that naturally produces the enzyme, using host organisms which
CC produce a relatively low level of expression. Therefore, industrial
CC production of the enzyme is not feasible using this method. This is a
CC much improved method which produces at least 250 mg enzyme per liter
CC of fermentation medium.
XX
SQ Sequence 546 AA;

Query Match 99.6%; Score 2985; DB 21; Length 546;
Best Local Similarity 99.5%; Pred. No. 5.7e-280;
Matches 543; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTAL 60

Qy         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DRAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYFVSSGDTNWG 120

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFSWDGFT 240

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHL 300

Qy        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGS 360

Qy        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEIHKVVW 420
              ||||||||||||||||||||||||||||||||||:|||||||||:||||||||||||||
Db        361 GPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTTAEMKDALLQLDMFGGEIHKVVW 420

Qy        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 DATAVAQREYIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQVESGKGV 480

Qy        481 FEGCYFNYPDVDLNNWKNGKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        481 FEGCYFNYPDVDLNNWKNAKYGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540

Qy        541 EPKQTK 546
              ||||||
Db        541 EPKQTK 546
```

RESULT 5

AAY39306
ID AAY39306 standard; Protein; 539 AA.
XX
AC AAY39306;
XX
DT 01-DEC-1999 (first entry)
XX
DE SpnJ protein involved in spinosyn biosynthesis.
XX
KW Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW PKS; extender module; initiator module; acyl transferase domain; AT;
KW acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW insecticide; oxido-reduction.
XX
OS Saccharopolyspora spinosa.
XX
PN WO9946387-A1.
XX
PD 16-SEP-1999.
XX
PF 16-FEB-1999; 99WO-US03212.
XX
PR 09-MAR-1998; 98US-0036987.
XX
PA (DOWC) DOW AGROSCIENCES LLC.
XX
PI Baltz RH, Broughton MC, Crawford KP, Madduri K, Merlo DJ;
PI Treadway PJ, Turner JR, Waldron C;
XX
DR WPI; 1999-551414/46.
DR N-PSDB; AAZ21501.
XX
PT New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT for production of insecticidal spinosyn compounds -
XX
PS Claim 1; Page 153-154; 190pp; English.
XX
CC This is the amino acid sequence of the product of the spnJ gene. The
CC protein is involved in spinosyn biosynthesis. The SpnJ gene is one of 23
CC genes and open reading frames contained in an 80kb DNA sequence
CC AAZ21501. Spinosyns are insecticidal microlides which are useful for the
CC control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC occurs via stepwise condensation and modification of carboxylic acid
CC precursors generating a linear polyketide which is modified further. The
CC DNA sequence contains a central region of approximately 55kb which has
CC homology to the DNA encoding the polyketide synthases (PKS) of known
CC macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC stop codons at the end of acyl carrier protein (ACP) domains. Together
CC the PKS polypeptides (AAY39297-Y39301), form a complex consisting of an
CC initiator module, spnA, and several extender modules spnB-spnE. The
CC products of the genes present in the region upstream of the PKS genes
CC have been assigned names spnF-spnS AAY39302-Y39315 and are responsible
CC for different modifications in spinosyn biosynthesis. There are also two
CC ORFs ORFL15 and ORFL16 present immediately upstream of spnS, producing
CC polypeptides AAY39316-Y39317, and two ORFs ORF1 and ORF2 present
CC downstream of the PKS region producing polypeptides AAY39318-Y39319. It
CC is suggested that SpnJ is involved in oxido reduction during spinosyn
CC biosynthesis. The genes are useful to improve yields of spinosyns, and
CC for creating new spinosyns e.g. by mutagenesis, or interruption of steps
CC in spinosyn biosynthesis. The modified spinosyns may be a new insect
CC control agent or serve as substrates for further chemical modification
CC and the creation of new semi-synthetic spinosyns. The genes are also
CC useful to isolate similar sequences from S. spinosa or other species by
CC hybridization.
XX
SQ Sequence 539 AA;

Query Match 21.2%; Score 635.5; DB 20; Length 539;
Best Local Similarity 30.2%; Pred. No. 2.5e-52;
Matches 174; Conservative 91; Mismatches 216; Indels 95; Gaps 16;

```
Qy          2 ATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTALD 61
              | : :: ||  :      | :  ||| | :  | | |: |   : :::  : :    |:
Db         25 APMNRRTPGTEI------TVEPDDPRYPDLVVGHNPRFTG-KPERIHIASSAEDVVHAVA 77

Qy         62 RAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYF-VSSGDTNWG 120
               |:   :   | : |||||:|: | |  :: ::::: |    ||  || | : :|
Db         78 DAVR--TGRRVGVRSGGHCFENLVADPAIRVLVDLSELNRVYYDSTRGAFAIEAGAALGQ 135

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ::|||:: |  :| |:|  || ||||:||| | |:|  |  ||:| ||||||    |
Db        136 VYRTLFKNWGVTIPTGACPGVGAGGHILGGGYGPLSRRFGSVVDYLQGVEVVVVDQAGEV 195

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKD----------LPMSPRGVIA 229
              :::   ::| |   :|:|||||||||||:|::: :             ||  |  |:
Db        196 HIVE-ADRNSTGAGHDLWWAHTGGGGGNFGIVTRFWLRTPDVVSTDAAELLPRPPATVLL 254

Qy        230 SNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSY----- 284
              : |: |   |  :   ||  :             | :   | | |   : |:::
Db        255 RSFHWPWHELTEQSFAVLLQNF-----------GNWYEQHSAPESTQLGLFSTLVCAHR 302

Qy        285 -------------SNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRP 330
                          :||| :|:    || |   |:  |  | : |     |
Db        303 QAGYVTLNVHLDGTDPNAERTLAE----HLSAINAQVGVT--PAEGLRETLPWL------ 350

Qy        331 RKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLT 390
              | |                |: |   |    || | |:||:    : |:  ::: ||
Db        351 --RSTQVAG-----------AIAE--GGEPGMQRTKVKAAYLRTGLSEAQLATVYRRLT 394

Qy        391 EVPDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNL 450
                :   :   | | :  :||  : |   |||:|||: ::|  : | | |  :|  :|
Db        395 -----VYGYDNPAAALLLLGYGGMANAVAPSATALAQRDSVLKALFVTNWSEPAEDERHL 449

Qy        451 KWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNN--WKNGKYGALELYF 508
              |||  || |||   |||| | |:|       :| | |||| || :  |        :||:
Db        450 TWIRGFYREMYAETGGVPVPGTRV-------DGSYINYPDTDLADPLWNTSGVAWHDLYY 502

Qy        509 LGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLKEPKQ 544
                |  || :||  |||  || :  ||       | |
Db        503 KDNYPRLQRAKARWDPQNIFQHGLSIKPPARLSPGQ 538
```

```
RESULT 6
AAB70952
ID   AAB70952 standard; Protein; 539 AA.
XX
AC   AAB70952;
XX
DT   28-AUG-2001  (first entry)
XX
DE   S. spinosa protein fragment encoded by ORF5, SEQ ID 16.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; C-C cyclising enzyme.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-1057268.
XX
PR   27-AUG-1999;   99DE-1040596.
XX
PA   (FARB ) BAYER AG.
XX
PI   Eberz G,  Moehrle V,  Froede R,  Velten R,  Salas JA;
XX
DR   WPI; 2001-267102/28.
DR   N-PSDB; AAF88322.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives
PT   -
XX
PS   Claim 33; Page 119-121; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence represents an S. spinosa C-C cyclising enzyme.
XX
SQ   Sequence   539 AA;

  Query Match             21.2%;  Score 635.5;  DB 22;  Length 539;
  Best Local Similarity   30.2%;  Pred. No. 2.5e-52;
  Matches  174;  Conservative   91;  Mismatches  216;  Indels   95;  Gaps   16;

Qy          2 ATLPQKDPGYIVIDVNAGTPDKPDPRLPSMKQGFNRRWIGTNIDFVYVVYTPQGACTALD 61
              | : :: ||  :      | :  ||| | :  | | |: |   : :::  : :    |:
Db         25 APMNRRTPGTEI------TVEPDDPRYPDLVVGHNPRFTG-KPERIHIASSAEDVVHAVA 77

Qy         62 RAMEKCSPGTVRIVSGGHCYEDFVFDECVKAIINVTGLVESGYDDDRGYF-VSSGDTNWG 120
               |:   :   | : |||||:|: | |  :: ::::: |    ||  || | : :|
Db         78 DAVR--TGRRVGVRSGGHCFENLVADPAIRVLVDLSELNRVYYDSTRGAFAIEAGAALGQ 135

Qy        121 SFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTED 180
              ::|||:: |  :| |:|  || ||||:||| | |:|  |  ||:| ||||||    |
Db        136 VYRTLFKNWGVTIPTGACPGVGAGGHILGGGYGPLSRRFGSVVDYLQGVEVVVVDQAGEV 195

Qy        181 SVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKD----------LPMSPRGVIA 229
              :::   ::| |   :|:|||||||||||:|::: :             ||  |  |:
Db        196 HIVE-ADRNSTGAGHDLWWAHTGGGGGNFGIVTRFWLRTPDVVSTDAAELLPRPPATVLL 254

Qy        230 SNLHFSWDGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSY----- 284
              : |: |   |  :   ||  :             | :   | | |   : |:::
Db        255 RSFHWPWHELTEQSFAVLLQNF-----------GNWYEQHSAPESTQLGLFSTLVCAHR 302

Qy        285 -------------SNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRP 330
                          :||| :|:    || |   |:  |  | : |     |
Db        303 QAGYVTLNVHLDGTDPNAERTLAE----HLSAINAQVGVT--PAEGLRETLPWL------ 350

Qy        331 RKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLT 390
              | |                |: |   |    || | |:||:    : |:  ::: ||
Db        351 --RSTQVAG-----------AIAE--GGEPGMQRTKVKAAYLRTGLSEAQLATVYRRLT 394

Qy        391 EVPDGLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVNL 450
                :   :   | | :  :||  : |   |||:|||: ::|  : | | |  :|  :|
Db        395 -----VYGYDNPAAALLLLGYGGMANAVAPSATALAQRDSVLKALFVTNWSEPAEDERHL 449

Qy        451 KWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNN--WKNGKYGALELYF 508
              |||  || |||   |||| | |:|       :| | |||| || :  |        :||:
Db        450 TWIRGFYREMYAETGGVPVPGTRV-------DGSYINYPDTDLADPLWNTSGVAWHDLYY 502

Qy        509 LGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLKEPKQ 544
                |  || :||  |||  || :  ||       | |
Db        503 KDNYPRLQRAKARWDPQNIFQHGLSIKPPARLSPGQ 538
```

```
RESULT 7
ABB91135
ID   ABB91135 standard; Protein; 530 AA.
XX
```

```
AC   ABB91135;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 346.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP09892.
XX
PR   28-AUG-2001; 2001WO-EP09892.
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K,  Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms -
XX
PS   Claim 5; SEQ ID NO 346; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence   530 AA;

  Query Match             12.3%;  Score 368.5;  DB 23;  Length 530;
  Best Local Similarity   27.0%;  Pred. No. 1.8e-26;
  Matches  134;  Conservative   76;  Mismatches  180;  Indels  107;  Gaps   23;

Qy          72 VRIVSGGHCYEDFVFDECVKAII-------NVTGLVESGYDDDRGYFVSSGDTNWGSFKT 124
               :|| |||| ||   :   |  :|       ::|  | |     :  :: :| |  |   |
Db         103 LRIRSGGHDYEGLSYMSSVPFVILDMYNLRSITVDVSS-----KKAWIQAGAT-LGELYT 156

Qy         125 LFRDHGRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSV 182
                 |  : |  | | | :|| |||| ||| | | | :|: || :   :::       |
Db         157 NVNDVSQTLAFPAGVCATVGAGGHISGGGYGNLMRKYGITVDHVIDAQII------DVN 209

Qy         183 LKYVHKDSEGNDGELFWAHTGGGGGNFGIIT--KYYFKDLPMSPRGVIASNLHFSWDGFT 240
               | ::: : | |  ||||  |||||:||:|   |    |:|      |   |       |
Db         210 GKLLNRATMGED--LFWAIRGGGGGSFGVILSWKINLVDVPKIVT-VFKVNKTLEQGG-- 264

Qy         241 RDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEE-----------FVMYLYTSYSND 287
                 |:| |         |:    ||   :| :|  :                :  |  :
Db         265 ----TDVLYK--------WQLVASKFPESLFVRAMPQVANGTKRGERTITVVFYAQFLGR 312

Qy         288 AEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMD 347
                :  :|       |  ::  :: |:    |      || :|          : || :    :|
Db         313 TDALMAIMNQNWPELGLK--HEDCQEMSWLNSTLFWADYPA------GTPTSIL----LD 360

Qy         348 YPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQ 407
                |           | |     | || |: |  |   :: :|| : :  : :          :|
Db         361 RP----------SSPGDFFKSKSDYVKKPIPKEGLEKLWKTMLKFNNNI-------VWMQ 403

Qy         408 VDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGG 466
                : :|| : ::   |||   |:  : |:|| | |   :    :|  ::: | |: |||
Db         404 FNPYGGVMDRIPATATAFPHRKGNLFKIQYFTTWFNANATMSSLSQMKELY-EVAEPY-- 460

Qy         467 VPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGK--------YGALELYFLGNLNRLIKA 518
                       | |         :||| |:|: :  :|:         ||:   |||||| ||:
Db         461 -------VSSNP---REAFFNYRDIDVGSNPSGETNVDEAKIYGS--KYFLGNLKRLMDV 508

Qy         519 KWLWDPNEIFTNKQSIP 535
                 |  :||:  | |:||||
Db         509 KAKYDPDNFFKNEQSIP 525


RESULT 8
ABB93022
ID   ABB93022 standard; Protein; 539 AA.
XX
AC   ABB93022;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 2233.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP09892.
XX
PR   28-AUG-2001; 2001WO-EP09892.
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K,  Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms -
XX
PS   Claim 5; SEQ ID NO 2233; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence   539 AA;

  Query Match             11.6%;  Score 347.5;  DB 23;  Length 539;
  Best Local Similarity   26.2%;  Pred. No. 2e-24;
  Matches  141;  Conservative   76;  Mismatches  181;  Indels  141;  Gaps   28;

Qy          52 TPQGACTALDR------AMEKCSPG---TVRIVSGGHCYE--DFVFDECVKAIINVTGLV 100
               ||: |     |       |   ||      ::| |||| ||   :: |:    |:::: |
Db          83 TPKPAIIVTPRSDIHVSAAVTCSKSLNFLLKIRSGGHDYEGLSYISDKPF-FILDMSNLR 141

Qy         101 ESGYD-DDRGYFVSSGDT-------NWGSFKTLFRDHGRVLPGGSCYSVGLGGHIVGGGD 152
               :   |  |:  ::|:| |         |    |      ||   | | | :||:|||| |||
Db         142 DVSVDIADQSAWISAGATLGEVYYRIWEKSKV----HG--FPAGVCPTVGVGGHISGGGY 195

Qy         153 GILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGII 212
               | : |  || || |   ::|       |   : : : | | |  |||| :|||| :||::
Db         196 GNMLRKFGLSVDNLIDAKIV-------DVNGQILDRKSMGED--LFWAISGGGGASFGVV 246
```

```
Qy        213 TKYYFKDLPMSPRGVI----------ASNLHFSWDG---------FTRDALQDLLTKYFK 253
Db        247 LGYKVKLVPVPETVTVFRVEKYMDSGAVDMVHKWQSVGPKTDRNLFLRMLIQPVTRKKVK 306

Qy        254 LARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEP 313
Db        307 TVRA----TV--VALFLGRAEEVVALLGKEFP--------------ELSLKK--ENCSE 343

Qy        314 TKALGGHAGW----APFPVRPRKRHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYK 369
Db        344 MTWFQSALWWDNRVNPTQIDPK--------VFLDRNLDRANF--------------GKRK 381

Qy        370 SAYMIKDFPDFQIDVIWKYLTEVPD-GLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQR 428
Db        382 SDYVASEIPRDGIESLFKKMTELGKIGLV----------FNPYGGKMAEVTVNATPFPHR 431

Qy        429 EYIIKLQYQTYWQEEDKDAVNLKWIRDFYEE---MYEPYGGVPDPNTQVESGKGVFEGCY 485
Db        432 SKLFKIQYSVTWQE---NSVEIE--KGFLNQANVLYSFMTGFVSKNPR---------NAY 477

Qy        486 FNYPDVDL-------NNWKNGK-YGALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPT 536
Db        478 LNYRDVDIGVNDHGTNSYEEGEVYG--RKYFGDNFDRLVKVKTAADPDNFFRNEQSIPT 534
```

RESULT 9
ABB91138
ID ABB91138 standard; Protein; 529 AA.
XX
AC ABB91138;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 349.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 349; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 529 AA;

Query Match 11.1%; Score 333.5; DB 23; Length 529;
Best Local Similarity 25.8%; Pred. No. 4.4e-23;
Matches 128; Conservative 77; Mismatches 184; Indels 107; Gaps 22;

```
Qy         72 VRIVSGGHCYEDFVFDECVKAII-------NVTGLVESGYDDDRGYFVSSGDTNWGSFKT 124
Db        103 LRIRSGGHDLEGLSYRSSVPFVILDMFNLRSITVNVLS-----KKAWVQAGAT-LGELYV 156

Qy        125 LFRDHGRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSV 182
Db        157 KINEASQTLAFPAGVCPTVGVGGHISGGGYGNLMRKFGITVDHVSDAQLI-------DVN 209

Qy        183 LKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPR--GVIASNLHFSWDGFT 240
Db        210 GKLLNRASMGED--LFWAIRGGGGASFGVILSWKI-NLVKVPKILTVFKVNKTLEQGG-- 264

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEEFV----------MYLYTSYSND 287
Db        265 ----TDVLYK--------WQLVATKFPEDLFMRAWPQIINGAERGDRTIAVVFYAQFLGP 312

Qy        288 AEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMD 347
Db        313 ADKLLAIMNQRLPELGLRR--EDCHEMSWFNTTLFWADYPAGTPK------SVLLDRPTN 364

Qy        348 YPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDAL-L 406
Db        365 PGFF---------------KSKSDYVKKPIPKEGLEKLWK---------TMFKFNNIVWM 400

Qy        407 QVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYG 465
Db        401 QFNPYGGVMDQIPSTATAFPHRKGNMFKVQYSTTWLAANATEISLSMMKELY-KVAEPY- 458

Qy        466 GVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGKYGALEL------YFLGNLNRLIKAK 519
Db        459 --------VSSNP---REAFFNYRDIDIGSNPSDETNVDEAKIYGYKYFLGNLKRLMQVK 507

Qy        520 WLWDPNEIFTNKQSIP 535
Db        508 AKYDPENFFKNEQSIP 523
```

RESULT 10
ABB91134
ID ABB91134 standard; Protein; 535 AA.
XX
AC ABB91134;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 345.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -

```
XX
PS   Claim 5; SEQ ID NO 345; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence   535 AA;

  Query Match             11.0%;  Score 328.5;  DB 23;  Length 535;
  Best Local Similarity   27.5%;  Pred. No. 1.4e-22;
  Matches  138;  Conservative   66;  Mismatches  180;  Indels  117;  Gaps   25;

Qy         72 VRIVSGGHCYEDFVFDECVKAII-------NVTGLVESGYDDDRGYFVSSGDTNWGSFKT 124
              :|| ||||  |   :   |  :|       ::|  | |     :  :| :| |  |
Db        103 IRIRSGGHDNEGLSYVSSVPFVILDMHKLRDITVDVSS-----KKAWVQAGAT-LGELYV 156

Qy        125 LFRDHGRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSV 182
                 :  : |  | | | :|| |||| ||| | | |  |  || :   |:|       |
Db        157 KIDEASQTLAFPAGICATVGAGGHISGGGYGNLMRKFGTTVDHVIDAELV-------DVN 209

Qy        183 LKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPR--GVIASNLHFSWDGFT 240
               | ::: : | |  ||||  |||| :||:|  :   :|   |:   |      |
Db        210 GKLLNRSTMGED--LFWAIRGGGGASFGVILSWKI-NLVEVPKIFTVFQVNKTLEQGG-- 264

Qy        241 RDALQDLLTKYFKLARCDWKNTVGKF--QIFHQAAEEFVMYLYTSYSNDAEREVA----- 293
                 |:: |         |:    ||   :| :|  : |     : :   || :|
Db        265 ----TDVVYK--------WQLVANKFPDNLFLRAMPQVV-----NGTKHGERTIAIVFWA 307

Qy        294 --QDRHYHLEADIEQIY-------KTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDE 344
                   |   : | :       : |:     |      ||  |        | ||  :
Db        308 QFLGRTDELMEIMNQSFPELGLRREDCQEMSWLNTTLFWAMLPA-----GTPKTVLLGRP 362

Qy        345 TMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDA 404
               | |  |:                | || |: |  |   :: |||          |  :  :
Db        363 T-DPVFF---------------KSKSDYVKKPIPKEGLEKIWK---------TMLKFNNI 397

Qy        405 L-LQVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYE 462
               : |  : :|| : ::  :|||    |:  : |:|| | | : :     ||  ::: | |: |
Db        398 VWLHFNPYGGMMDRIPSNATAFPHRKGNLFKVQYYTTWLDPNATESNLSIMKELY-EVAE 456

Qy        463 PYGGVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGK--------YGALELYFLGNLNR 514
               ||          | |         :||| |:|: :  :|:          ||    ||||| |
Db        457 PY---------VSSNP---REAFFNYRDIDIGSNPSGETDVDEAKIYG--YKYFLGNLKR 502

Qy        515 LIKAKWLWDPNEIFTNKQSIP 535
               |:  |   ||   | |:||||
Db        503 LMDVKAKSDPENFFKNEQSIP 523


RESULT 11
ABB90805
ID   ABB90805 standard; Protein; 541 AA.
XX
AC   ABB90805;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 16.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP09892.
XX
PR   28-AUG-2001; 2001WO-EP09892.
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K,  Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms -
XX
PS   Claim 5; SEQ ID NO 16; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence   541 AA;

  Query Match             10.8%;  Score 324.5;  DB 23;  Length 541;
  Best Local Similarity   25.4%;  Pred. No. 3.4e-22;
  Matches  132;  Conservative   73;  Mismatches  169;  Indels  145;  Gaps   26;

Qy         76 SGGHCYED---------FVFDECVKAIINVTGLVESGYDDDRGYFVSSGDT-------NW 119
              |||| |:           || |  :  : |:|  |:   ||    :| :| |        |
Db        110 SGGHDYDGVSYISNRPFFVLD--MSYLRNIT--VDMS-DDGGSAWVGAGATLGEVYYNIW 164

Qy        120 GSFKTLFRDHG-RVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLT 178
               | ||    ||    | | | :|| |||| ||| | : | :|| ||:::  ::|
Db        165 QSSKT----HGTHGFPAGVCPTVGAGGHISGGGYGNMIRKYGLSVDYVTDAKIV------ 214

Qy        179 EDSVLKYVHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVI---------- 228
               |   : : : | | |  ||||  |||| :||:|  :  | :|: ||  :
Db        215 -DVNGRILDRKSMGED--LFWAIGGGGGASFGVILSFKIKLVPVPPRVTVFRVEKTLVEN 271

Qy        229 ASNLHFSW---------DGFTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMY 279
               | ::   |         | | |  || :            :||       |     |:
Db        272 ALDMVHKWQFVAPKTSPDLFMRLMLQPVT----------RNTT-------QTVRASVVA 313

Qy        280 LYTSYSNDAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWA----PFPVRPR---K 332
               |:    :|    : ::     |  ::   : |      :      ||       ::|
Db        314 LFLGKQSDLMSLLTKE---FPELGLKP--ENCTEMTWIQSVMWWANNDNATVIKPEILLD 368

Qy        333 RHTSKTSYMHDETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEV 392
               |:    |::                          | || |: |:     :| : | | |
Db        369 RNPDSASFL-----------------------KRKSDYVEKEISKDGLDFLCKKLMEA 403

Qy        393 PD-GLTSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEEDKDAVN-- 449
                 ||             : :||:: :|   ||     |: : |:|:    |::   |  :
Db        404 GKLGLV----------FNPYGGKMSEVATTATPFPHRKRLFKVQHSMNWKDPGTDVESSF 453

Qy        450 LKWIRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDL-------NNWKNGK-Y 501
               ::  | ||  |  |:      :|           | || |:|:       |:::  : |
Db        454 MEKTRSFYSYM-APFVTKNPRHT-----------YLNYRDLDIGINSHGPNSYREAEVY 500

Qy        502 GALELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPTKPLK 540
               |    ||  | :||:| |   ||    | ::|||| | |
Db        501 G--RKYFGENFDRLVKVKTAVDPENFFRDEQSIPTLPTK 537


RESULT 12
ABB93021
```

```
ID   ABB93021 standard; Protein; 540 AA.
XX
AC   ABB93021;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 2232.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP09892.
XX
PR   28-AUG-2001; 2001WO-EP09892.
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K,  Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms -
XX
PS   Claim 5; SEQ ID NO 2232; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence   540 AA;

  Query Match            10.7%;  Score 322;  DB 23;  Length 540;
  Best Local Similarity   24.4%;  Pred. No. 6e-22;
  Matches  125;  Conservative   79;  Mismatches  172;  Indels  136;  Gaps   22;

Qy         72 VRIVSGGHCYE--DFVFDECVKAIINVTGLVESGYD-DDRGYFVSSGDTNWGSFKTLF--  126
              ::| |||| |:   :: |:    |:::: : :   |      ::|:| |    :  ::
Db        113 LKIRSGGHDYDGLSYISDKPF-FILDMSNIRDVSVDIASNSAWISAGATLGEVYYRIWEK  171

Qy        127 -RDHGRVLPGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKY  185
               | ||   | | | :||:|||: |||  | : |  || ||::   ::|       |   :
Db        172 SRVHG--FPAGVCPTVGVGGHLSGGGYGNMVRKFGLSVDYVEDAKIV-------DVNGRV  222

Qy        186 VHKDSEGNDGELFWAHTGGGGGNFGIITKYYFKDLPMSPRGVI----------ASNLHFS  235
              : : :| |  |||| ||||||::|::  |  | :|:     :           | ::
Db        223 LDRKAMGED--LFWAITGGGGGSYGVVLGYKVKLVPVPSVVTVFRVEQYMDSGAVDMVHK  280

Qy        236 WDG----------FTRDALQDLLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSN  286
              |          | |  :| :  |  |  |           :|   |:| |  |   :
Db        281 WQSVGPKTDPNLFMRMLIQPVTRKKVKTVR------ASVVALFLGRADEVVALLSKEF--  332

Qy        287 DAEREVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMH----  342
                                          |   |             :| : |: ::
Db        333 -------------------------PELGL------------KKENCSEMTWFQSALW  353

Qy        343 -----DETMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPD-GL  396
                   : |   |   |   :: |     || || |:     |   |: ::| : |:   ||
Db        354 WDNRLNATQVDPKVFLDRNLDTS---SFGKRKSDYVATAIPKKGIESLFKKMIELGKIGL  410

Qy        397 TSAEMKDALLQVDMFGGEIHKVVWDATAVAQREYIIKLQYQTYWQEE----DKDAVNLKW  452
                          : :||:: :|  :|     |  : |:||   |:|     :|  :|
Db        411 V----------FNPYGGKMAEVAVNAKPFPHRNKLFKIQYSVNWKENSAEIEKGYLNQAK  460

Qy        453 IRDFYEEMYEPYGGVPDPNTQVESGKGVFEGCYFNYPDVDL-------NNWKNGK-YGAL  504
              :     :|    |    | :           ||||  |||:       |::| |: ||
Db        461 V------LYSFMTGFVSKNPR---------SSYFNYRDVDIGVNDHGANSYKEGEVYG--  503

Qy        505 ELYFLGNLNRLIKAKWLWDPNEIFTNKQSIPT  536
                ||  | :||:| |   ||   | |:|||||
Db        504 RKYFGENFDRLVKIKTAVDPGNFFRNEQSIPT  535


RESULT 13
AAY88391
ID   AAY88391 standard; Protein; 545 AA.
XX
AC   AAY88391;
XX
DT   27-JUL-2000  (first entry)
XX
DE   Tetrahydrocannabinolic acid synthase amino acid sequence.
XX
KW   Tetrahydrocannabinolic acid synthase; THCA; anaesthesia; pain-killer;
KW   intraocular pressure lowering; anti-inflammatory treatment.
XX
OS   Cannabis sativa.
XX
PN   JP2000078979-A.
XX
PD   21-MAR-2000.
XX
PF   04-SEP-1998;   98JP-0251667.
XX
PR   04-SEP-1998;   98JP-0251667.
XX
PA   (TAIS ) TAISHO PHARM CO LTD.
XX
DR   WPI; 2000-285930/25.
DR   N-PSDB; AAA13340, AAA13341.
XX
PT   New tetrahydrocannabinolic acid synthase gene  -
XX
PS   Claim 2; Page 8-10; 17pp; Japanese.
XX
CC   This sequence represents the Cannabis sativa tetrahydrocannabinolic acid
CC   (THCA) synthase amino acid sequence. The THCA gene sequence can be used
CC   to produce an expression vector containing the gene. Cells can be
CC   transformed using the vector, so that they produce a protein with THCA
CC   synthase activity. THCA can be used for anaesthesia, pain-killing,
CC   intraocular pressure lowering and anti-inflammatory treatment.
XX
SQ   Sequence   545 AA;

  Query Match            10.2%;  Score 306.5;  DB 21;  Length 545;
  Best Local Similarity   24.6%;  Pred. No. 1.9e-20;
  Matches  139;  Conservative   65;  Mismatches  207;  Indels  153;  Gaps   25;

Qy         36 NRRWIGTNIDFVYVVYTPQGACTALDRAMEKCSPGT---VRIVSGGHCYEDFVFDECVKA  92
              | |:|         |: ||     :  :|   ||       :|  ||||  |   :  |
Db         70 NLRFISDTTPKPLVIVTPSN--NSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQVPF  127

Qy         93 II--------------NVTGLVESGYDDDRGYFVSSGDTNWGSFKTLFRDHGRVLPGGSC  138
              ::               : |  ||:|       |:       | : |    :     ||| |
Db        128 VVVDLRNMHSIKIDVHSQTAWVEAGATLGEVYY-------WINEK----NENLSFPGGYC  176

Qy        139 YSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELF  198
              :||:|||  ||| | | | :||  |      ::   ::  |   | : : | | |  ||
Db        177 PTVGVGGHFSGGGYGALMRNYGLAAD-----NIIDAHLVNVDG--KVLDRKSMGED--LF  227

Qy        199 WAHTGGGGGNFGIITKYYFKDLPMSPRGVIAS----------NLHFSWDGFTRDALQDL  247
              ||  |||| |||||  :  | : :  :  | |          |   |        :||
Db        228 WAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDL  287
```

```
Qy        248 ------LTKYFKLARCDWKNTV-GKF-QIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYH 299
                    :||         | || | |  |||   :  |  :  |
Db        288 VLMTHFITKNITDNHGKNKTTVHGYFSSIFHGGVDSLVDLMNKS---------------- 331

Qy        300 LEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTETING 359
                                      ||    |:   |     | |:   ||:     |
Db        332 -------------------------FPELGIKKTDCKEFSWIDTTI---FYSGVVNFNT 362

Qy        360 SG-------PNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDM-- 410
                  :          || |:|: ||      :| : | : |          :: : | : |:
Db        363 ANFKKEILLDRSAGK-KTAFSIK------LDYVKKPIPET----AMVKILEKLYEEDVGA 411

Qy        411 -------FGGEIHKVVWDATAVAQREYII-KLQYQTYWQEEDKDAVNLKWIRDFYEEMYE 462
                    :|| : ::   |      |  |: :| |   |:::: :  :: |:|  |
Db        412 GMYVLYPYGGIMEEISESAIPFPHRAGIMYELWYTASWEKQEDNEKHINWVRSVY-NFTT 470

Qy        463 PYGGVPDPNTQVESGKGVFEGCYFNYPDVDL--------NNWKNGKYGALELYFLGNLNR 514
                  ||      | ::          | || |:||         ||:   :    | ||  | ||
Db        471 PY---VSQWPRL---------AYLNYRDLDLGKTNHASPNNYTQARIWG-EKYFGKNFNR 517

Qy        515 LIKAKWLWDPNEIFTNKQSIPTKP 538
                |:| |   |||  | |:||||  |
Db        518 LVKVKTKVDPNNFFRNEQSIPPLP 541
```

RESULT 14
ABB91137
ID ABB91137 standard; Protein; 552 AA.
XX
AC ABB91137;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 348.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 348; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 552 AA;

Query Match 10.1%; Score 302.5; DB 23; Length 552;
Best Local Similarity 25.3%; Pred. No. 4.8e-20;
Matches 128; Conservative 72; Mismatches 190; Indels 115; Gaps 24;

```
Qy         72 VRIVSGGHCYEDFVFDECVK-AIINVTGLVESGYDDDR-GYFVSSGDTNWGSFKTLFRDH 129
              :|| ||||  | | :   |   |:::  |     :  |  :| :| |  |       :
Db        127 LRIRSGGHDNEGFSYMSSVPFVILDMHNLRSIDVNLSRKNAWVQAGAT-LGELYVKINEA 185

Qy        130 GRVL--PGGSCYSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVH 187
               : |  | | | :|| |||| ||| | | |  |: || :   :::       |   | ::
Db        186 SQTLAFPAGVCPTVGAGGHISGGGFGNLMRKFGITVDHVIDAQII-------DVNGKLLN 238

Qy        188 KDSEGNDGELFWAHTGGGGGNFGIITKY-----------YFKDLPMSPRGVIASNLHFS 235
              : : | |  |||   ||| :||:|  :            ||      :|   ::: :
Db        239 RAAMGED--LFWA-IRGGGSSFGVILSWKINLVEVPKILTVFKVNKTLEQG--GTDILYK 293

Qy        236 WDGFTRDALQDLLTKYFKLARCDWKNTV-----GKFQIFHQAAEEFVMYLYTSYSNDAER 290
              |    : | | |  | |   | ||     |: :          | |       |
Db        294 WQ-LVANKLPDSL---FITA---WPRTVNGPKPGERTV------AVVFYAQFLGPTDKLM 340

Qy        291 EVAQDRHYHLEADIEQIYKTCEPTKALGGHAGWAPFPVRPRKR------HTSKTSYMHDE 344
              |:      |    |     |      || :|    |        |:  |:
Db        341 EIMDQSFPELGLGRED----CHEMSWLNTTLFWANYPAGTPKSILLDRPPTNSVSF---- 392

Qy        345 TMDYPFYALTETINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDA 404
                                 | || :: |  |   :: :|| : :    ::
Db        393 ---------------------KSKSDFVKKPIPKKGLEKLWKTMFKFNSSVS------- 423

Qy        405 LLQVDMFGGEIHKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEP 463
               || : :|| : ::   |||    |:  : |:|| | | : :    :|  : : : |: ||
Db        424 -LQFNPYGGVMDRIPATATAFPHRKGNLFKVQYSTMWFDANATESSLAMMNELF-EVAEP 481

Qy        464 YGGVPDPNTQVESGKGVFEGCYFNYPDVDLNNWKNGK--------YGALELYFLGNLNRL 515
              |         | |         :||: |:|: :  :|:         ||:   ||||| ||
Db        482 Y---------VSSNP---REAFFNFRDIDIGSNPSGETNVDEAKIYGS--KYFLGNLKRL 527

Qy        516 IKAKWLWDPNEIFTNKQSIPTKPLK 540
               :  |  :||:  | |:||||   :|
Db        528 MDVKAKYDPDNFFKNEQSIPPVRVK 552
```

RESULT 15
ABB93020
ID ABB93020 standard; Protein; 532 AA.
XX
AC ABB93020;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2231.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -

XX
PS Claim 5; SEQ ID NO 2231; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 532 AA;

Query Match 10.0%; Score 300; DB 23; Length 532;
Best Local Similarity 23.5%; Pred. No. 7.9e-20;
Matches 129; Conservative 78; Mismatches 196; Indels 146; Gaps 24;

```
Qy          36 NRRWIGTNIDFVYVVYTPQGACTALDRAMEKCSPG---TVRIVSGGHCYE--------DF 84
               |:|:   ::    :: ||        ::  ||:      :|  |||| ||         |
Db          73 NQRFSAPDVPKPVLILTPVQPSDV--QSAVKCARRFGIHIRTRSGGHDYEGLSYVTHKPF 130

Qy          85 VFDECVKAIINVTGLVESGYD-DDRGYFVSSGDTNWGSFKTLFRDHGR-----VLPGGSC 138
               |       |:::  |      | |:|  :| :| |       |: : |:       | | |
Db         131 V-------ILDLRNLRSITVDVDNRSVWVQTGAT----IGELYYEIGKKNRTLAFPAGVC 179

Qy         139 YSVGLGGHIVGGGDGILARLHGLPVDWLSGVEVVVKPVLTEDSVLKYVHKDSEGNDGELF 198
                :||:|||  ||| | | | |||  | :    ||       |:  : : :   | |   |
Db         180 PTVGVGGHFSGGGYGTLLRKHGLAADHVIDARVV-------DARGRILERREMGED--FF 230

Qy         199 WAHTGGGGGNFGIITKYYFKDLPMSPRGVIASNLHFS----------WDGFTRDALQD-- 246
               ||  |||| :| ::  :    : :     : :   ||          |  | | : |
Db         231 WAIRGGGGSSFCVVLSWKIGLINVPSTVTVFNVTKFSEQSALKIIHRWQ-FVADKVSDDL 289

Qy         247 ----LLTKYFKLARCDWKNTVGKFQIFHQAAEEFVMYLYTSYSNDAEREVAQDRHYHLEA 302
                   :| :|  : |  :                    :||  |  |  :    :     ||
Db         290 FIRVMLQRYKNMVRASFPG---------------LYL-GSVKNLLKMVNKEFPELGLEE 332

Qy         303 D-------IEQIYKTCEPTKALGGHAGWAPFPVRPRKRHTSKTSYMHDETMDYPFYALTE 355
               |       || :    |    ||                                : |  ||:
Db         333 DDCTEMSWIESVIWFAE----LG----------------------------EEPINVLTK 360

Qy         356 TINGSGPNQRGKYKSAYMIKDFPDFQIDVIWKYLTEVPDGLTSAEMKDALLQVDMFGGEI 415
                    |       | || :: :  |    |  :|: | |         | : | |    |||::
Db         361 RTRAS---LAFKAKSDFVQEPMPKTAISKLWRRLQE-------PEAEHAQLIFTPFGGKM 410

Qy         416 HKVVWDATAVAQRE-YIIKLQYQTYWQEEDKDAVNLKWIRDFYEEMYEPYGGVPDPNTQV 474
                ::    |    |:  | ::||  ||: : |:   ::|:   |::| |     |
Db         411 SEIADYETPFPHRKGNIYEIQYLNYWRGDVKEKY-MRWVERVYDDMSEFVAKSP------ 463

Qy         475 ESGKGVFEGCYFNYPDVDL--------NNWKNGKYGALELYFLGNLNRLIKAKWLWDPNE 526
                      | | |  |:||        : :: ||   :: ||  |  ||:: |   ||::
Db         464 -------RGAYINLRDLDLGMYVGVKRSKYEEGKSWGVK-YFKNNFERLVRVKTSVDPSD 515

Qy         527 IFTNKQSIP 535
                | ::||||
Db         516 FFCDEQSIP 524
```

Search completed: June 10, 2003, 09:58:44
Job time : 45 secs